

Научная статья
УДК 94(47):330.15
<https://doi.org/10.23859/2587-8344-2023-7-2-1>

Природно-географическая среда как фактор накопления человеческого капитала в Европейской России (конец XIX в.)

Дмитрий Валерьевич Диденко✉

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации,
Москва, Россия,
didenko-dv@ranepa.ru✉, <http://orcid.org/0000-0001-5295-2538>

Dmitrii V. Didenko✉

Russian Presidential Academy of National Economy
and Public Administration,
Moscow, Russia,
didenko-dv@ranepa.ru✉, <https://orcid.org/0000-0001-5295-2538>



Евгений Сергеевич Гришин

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации,
Москва, Россия,
bibliosof-info@yandex.ru, <https://orcid.org/0000-0002-9521-2246>

Evgenii S. Grishin

Russian Presidential Academy of National Economy
and Public Administration,
Moscow, Russia,
bibliosof-info@yandex.ru, <https://orcid.org/0000-0002-9521-2246>



Аннотация. Целью данной статьи является ознакомление с результатами проведенного исследования, посвященного картографической визуализации и статистической проверке гипотезы о значимости природно-географических факторов в формировании человеческого капитала поздней Российской империи и количественной оценке их роли. В качестве косвенного показателя накопленного человеческого капитала используется уровень грамотности. Анализировались собранные в результате Первой всеобщей переписи населения 1897 г. демографические сведения, характеристики природно-географической среды, а также показатели развития транспортной инфраструктуры и аграрной

© Диденко Д.В., Гришин Е.С., 2023

© Didenko D., Grishin E., 2023

специализации уездной экономики. Единицей анализа выбран уезд европейской части страны (без Финляндии, Польши, Закавказья). Применялись методы корреляционного, регрессионного анализа и картографического исследования. Полученные результаты показывают, что ряд существенных факторов природной среды значимо ($p < 0,1$) связан с пространственной дифференциацией накопленного к 1897 г. уровня человеческого капитала Европейской России и его диспаритетом в разрезе «город/село». В меньшей степени факторы природной среды связаны с диспаритетом в гендерном разрезе.

Ключевые слова: природная среда, климат, региональная экономика, технологии, Россия позднеимперского периода, институты, социально-экономическое неравенство

Благодарность. Статья подготовлена в рамках выполнения НИР ФГБОУ ВО РАНХиГС № 10.20-2023-2 «Факторы социально-экономического развития регионов Российской империи в XIX – начале XX вв.»

Авторы выражают благодарность Т.Я. Валетову и А.И. Дегтяревой за предоставленные в формате электронных таблиц демографические сведения.

Для цитирования: Диденко Д.В., Гришин Е.С. Природно-географическая среда как фактор накопления человеческого капитала в Европейской России (конец XIX в.) // *Historia provinciae* – журнал региональной истории. 2023. Т. 7, № 2. С. 381–436. <https://doi.org/10.23859/2587-8344-2023-7-2-1>

Natural and Geographical Environment as a Factor of Human Capital Accumulation in European Russia (Late 19th Century)

Abstract. The purpose of this article is to introduce the results of the study on cartographic visualization and statistical verification of the hypothesis about the significance of natural and geographical factors in the formation of human capital in the late Russian Empire and the quantitative assessment of their role. The level of literacy is used as an indirect (proxy) indicator of accumulated human capital. The demographic information collected as a result of the First General Population Census in 1897, the characteristics of the natural and geographical environment, as well as indicators of the development of transport infrastructure and agricultural specialization of the county (uyezd) economy were analyzed. The unit of analysis is an uyezd in the European part of the country (without Finland, Poland, and Transcaucasia). The methods of correlation, regression analysis, and cartographic research were applied. The results obtained show that a number of essential factors of the natural environment are significantly ($p < 0.1$) associated with the spatial differentiation of the level of human capital accumulated by 1897 in European Russia and its disparity in the urban/rural context. To a lesser extent, environmental factors are associated with gender disparity.

Keywords: natural environment, climate, regional economy, technologies, late imperial Russia, institutions, socio-economic inequality

Acknowledgement. The article was prepared as part of the research work of the Federal State Budgetary Educational Institution of Higher Education RANEPА no. 10.20-2023-2 “Factors of Socio-Economic Development of the Regions of the Russian Empire in the 19th – Early 20th Centuries.”

The authors are grateful to T. Valetov and A. Degtyareva for providing demographic information in spreadsheet format.

For citation: Didenko, D.V., and E.S. Grishin. “Natural and Geographical Environment as a Factor of Human Capital Accumulation in European Russia (Late 19th Century).” *Historia*

Provinciae – the Journal of Regional History, vol. 7, no. 2 (2023): 381–436, <https://doi.org/10.23859/2587-8344-2023-7-2-1>

Введение

Традиции объяснения процессов социально-экономического развития влиянием таких факторов, как климат и другие физико-географические особенности территорий, не только широко популярны, но и подтверждаются многими эмпирическими исследованиями, преимущественно на материале зарубежных стран. В работах по истории России данная проблематика осмысливается как влияние природно-климатического и географического факторов на ход отечественной истории.

Согласно Р. Пайпсу¹, природные условия способствовали генезису системы институтов «вотчинного государства» (отношений «власти-собственности» в терминологии постмарксистского институционализма Р.М. Нуреева, Ю.В. Латова²). Аналогичный взгляд, во многом основанный на идеях В.О. Ключевского о решающей роли географических факторов в российском историческом процессе, в условиях преобладания аграрной экономики, был представлен Л.В. Миловым³.

П.В. Турчин и С.А. Нефедов актуализировали понятия «емкость экологической ниши» и «несущая способность окружающей среды», которые отражают исторически определенный уровень взаимодействия ресурсов и технологий⁴. В структурно-демографической теории Дж. Голдстоуна⁵ указанные понятия операционализировались такими показателями, как численность населения (и отдельно элиты) по отношению к ресурсам, отношение фактической численности населения к максимально обеспечиваемой в соответствии с прожиточным минимумом при имеющихся технологиях (уровень давления населения на ресурсы)⁶.

¹ *Пайпс Р.* Россия при старом режиме / перевод с английского В. Козловского. Москва: Независимая газета, 1993. С. 12–41.

² *Нуреев Р.М., Латов Ю.В.* Экономическая история России (опыт институционального анализа). Москва: КНОРУС, 2016.

³ *Милов Л.В.* Великорусский пахарь и особенности российского исторического процесса. Москва: РОССПЭН, 2001. С. 4–17, 31–38, 383–385.

⁴ *Turchin P., Nefedov S.A.* Secular cycles. Princeton, NJ: Princeton University Press, 2009. P. 6, 18.

⁵ *Goldstone J.A.* Revolution and Rebellion in the Early Modern World. Berkeley: University of California Press, 1991.

⁶ *Turchin P., Nefedov S.A.* Secular cycles. P. 81.

Б.Н. Миронов привел ряд эмпирических свидетельств, опровергающих гипотезы, которые сформулированы на основе географического детерминизма, настаивающего на решающей роли климата в истории России⁷.

Р.М. Нуреев и Ю.В. Латов в значительной степени использовали количественные свидетельства и ряд концептуальных идей Р. Пайпса и Л.В. Милова, однако выводы о значимости природно-географических и институциональных факторов, к которым пришли исследователи, оказались ближе к выводам Б.Н. Миронова. Они проводили мысль, что природная среда являлась весьма существенной составляющей, значительно влияла на результаты конкуренции институтов при исторически заданных альтернативных вариантах (в большей степени в ранний период, чем в XX в.), но вряд ли когда-либо определяла ее исход⁸.

Указанные работы по преимуществу содержали нарратив. Количественные свидетельства по рассматриваемой проблеме, без обращения к эконометрическому инструментарию, в наибольшей степени содержались в работах Б.Н. Миронова, П.В. Турчина и С.А. Нефедова.

Среди работ, основанных на применении эконометрических методов, отметим статью Й. Батена и Р. Хиппе⁹, которые при тестировании моделей влияния природной среды на неравенство земельной собственности¹⁰ использовали материалы по Европейской России рассматриваемого нами периода.

Цель нашего исследования – картографически визуализировать и статистически проверить гипотезу о значимом влиянии природно-географических факторов на формирование человеческого капитала поздней Российской империи, приведя их количественные оценки.

Характеристика данных источников

В ходе работы авторы анализировали несколько групп показателей. В качестве ключевых рассматривались демографические сведения, собранные в

⁷ *Миронов Б.Н.* Российская империя: от традиции к модерну: в 3 т. Санкт-Петербург: ДМИТРИЙ БУЛАНИН, 2018. Т. 1. С. 221–250.

⁸ *Нуреев Р.М., Латов Ю.В.* Экономическая история России (опыт институционального анализа). С. 37–46.

⁹ *Baten J., Hippe R.* Geography, land inequality and regional numeracy in Europe in historical perspective // *Journal of Economic Growth*. 2018. Vol. 23, № 1. P. 79–109.

¹⁰ *Sokoloff K.L., Engerman S.L.* History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World // *Journal of Economic Perspectives*. 2000. Vol. 14, № 3. P. 217–232; *Engerman S.L., Sokoloff K.L.* Factor Endowments, Inequality, and Paths of Development among New World Economies // *Economía Journal*. 2002. Vol. 3, № 1. P. 41–109; *Galor O., Moav O., Vollrath D.* Land Inequality and the Emergence of Human Capital Promoting Institutions // *Review of Economic Studies*. 2009. Vol. 76, № 1. P. 143–179.

ходе Первой всеобщей переписи населения в 1897 г.¹¹: численность населения и количество грамотных в разрезах «мужчины/женщины» и «город/село» в возрастных когортах 20–29 и 40–49 лет¹². Таким образом, накопленный человеческий капитал косвенно характеризуется долей грамотных (в данном случае обладающих способностью читать на родном языке).

Другая группа – характеристики природной среды. Это физико-географические показатели: площадь территории, долгота, широта, геоморфологическое и географическое положение относительно водных путей, уровень плотности водных потоков и флювиальных форм рельефа¹³, почвенные условия¹⁴, преобладающий угол наклона рельефа¹⁵. К ним примыкают климатические показатели: количество дней с осадками за год¹⁶ и расчетная по ежемесячным данным среднегодовая температура¹⁷.

Геоморфологическая характеристика положения относительно водных путей включает следующие категории:

¹¹ Первая всеобщая перепись населения Российской империи 1897 г. / под редакцией Н.А. Тройницкого: в 89 т. Санкт-Петербург: Центральный статистический комитет Министерства внутренних дел, 1897–1905. Т. 1–50, 65, 67, 68, 70.

¹² Предоставлены в формате электронных таблиц Т.Я. Валетовым как научным руководителем бакалаврской ВКР А.И. Дегтяревой, 2021 г.

¹³ Величина характеризует отношение площади зоны к общему количеству различных форм обводненности и флювиальных неровностей. Вычислена на основе обработки в цифровых ГИС-системах источника: *Стрельбицкий И.А.* Специальная карта Европейской России (1:420 000). 1865–1871 гг. Петроград: Литография картографического заведения Военно-топографического отдела, 1915. Выделено 8 модульных групп по мере возрастания плотности.

¹⁴ См.: Статистическое обозрение Российской империи. Территория / составитель Л.Ф. де Ливрон. Санкт-Петербург: Общественная польза, 1874.

¹⁵ См.: Гипсометрическая карта Европейской России: опыт изображения строения поверхности Европейской России / составитель А.А. Тилло. Санкт-Петербург: Департамент шоссейных и водных сообщений Министерства путей сообщения, 1889. Установлен на основе ЦРМ после преобразования исходного растра в растр со значениями наибольших уклонов. Полученные величины были сгруппированы в 9 модульных групп.

¹⁶ Атлас распределения атмосферных осадков на речных бассейнах Европейской России по месяцам и за весь год на основании двадцатилетних наблюдений 1871–1890 / составитель А.А. Тилло. Санкт-Петербург: Типо-литография К. Биркенфельда, 1897; *Небольсин С.И.* Атлас карт среднего распределения атмосферных осадков в Европейской России. За отдельные месяцы и за год по наблюдениям 1888–1912 гг. Петроград: Типография «К. Биркенфельд», 1916.

¹⁷ См.: О температуре воздуха в Российской империи, на основании критического рассмотрения наблюдений, по предложению министра государственных имуществ П.А. Валуева, разработал директор Главной физической обсерватории и член Императорской Академии наук Г. Вильд: в 2 вып. Санкт-Петербург: Типография Императорской Академии наук, 1878–1882. Вып. 2. Приложение.

- 1 – коренной берег;
- 2 – пойменный берег;
- 3 – отсутствие выраженного характера асимметрии местности.

Характер географического положения (*geo_position*) населенного пункта идентифицируется следующими категориями:

- 1 – речная долина;
- 2 – эстуарий, дельта;
- 3 – устье при слиянии рек;
- 4 – озерный берег;
- 5 – морское побережье.

В Европейской части России выбрано 13 географических зон в зависимости от характеристики почв:

- 1 – Черноземная нестепная полоса;
- 2 – Черноземная степная полоса;
- 3 – Черноземно-суглинистая область;
- 4 – Степная нечерноземная область;
- 5 – Нечерноземная область с перегноем;
- 6 – Нечерноземная малоплодородная область;
- 7 – Пинское полесье, малоплодородная почва, леса и болота;
- 8 – Северо-Белорусское полесье, 35 % лесов, более плодородна, чем Пинское Полесье;
- 9 – Новгородско-Финское полесье, преобладание болот и лесов, неплодородная почва;
- 10 – Вологодско-Вятское полесье, преобладание лесов, умеренное плодородие почвы;
- 11 – Уральское полесье, малоплодородная почва, 75 % лесов;
- 12 – Крайний Север, сплошные леса, ограниченное плодородие почвы;
- 13 – Подгорная и нагорные области (без детальной характеристики почв).

Показателем развития железнодорожной транспортной инфраструктуры была определена продолжительность ее функционирования, а именно – время от открытия станции до переписи 1897 г.¹⁸

Отраслевую специализацию местной экономики (а именно роль производства зерна) характеризует чистый сбор зерновых на душу населения¹⁹.

¹⁸ См.: Краткие сведения о развитии отечественных железных дорог с 1838 г. по 1990 г. / составитель Г.М. Афонина. Москва: Издательство Центральной научно-технической библиотеки железнодорожного транспорта России, 1995; *Скорино И.А.* Указатель русских железных дорог. Санкт-Петербург: Военная типография, 1887.

¹⁹ Средний сбор хлебов и картофеля за десятилетие 1883–92 гг. в 60 губерниях Европейской России по отношению к народному продовольствию // *Временник*

Единицей анализа выступал уезд. В выборку европейской части страны не вошли территории Финляндии, Польши, Закавказья, которые к 1897 г. находились в составе Российской империи менее 100 лет и были в наименьшей степени интегрированы в социально-экономическое пространство страны.

В «среднюю выборку» вошел 421 уезд (в наличии данные по большинству показателей: площадь, координаты, производство зерна на душу населения и геологические условия, без почв); в «малую выборку» – 59 уездов (дополнительно имелись данные по климату). В «среднюю выборку с усечением» вошли 379 уездов, по которым, дополнительно к данным показателей полной «средней выборки», имелись данные по категории почв.

Методы анализа

Для реализации поставленной цели применялись методы корреляционного, регрессионного анализа и картографического исследования. Первый выступал предпосылкой второго.

В качестве зависимых переменных последовательно принимались 1 прокси-индикатор *уровня* человеческого капитала (доля грамотных обоого пола в возрасте 20–29 и 40–49 лет²⁰ в уезде – L) (см. картографическое отображение на рис. П1) и 2 показателя, характеризующие его *внутритерриториальное неравенство* (диспаритеты «город/село» – SG и «мужчины/женщины» – GG). Диспаритет в грамотности в разрезе «город/село» вычислялся как отношение доли грамотных в городах к доле грамотных в сельской местности. Гендерный диспаритет в грамотности вычислялся как отношение доли грамотного мужского населения к доле грамотного женского населения. Оба диспаритета имеют картографические отображения на рис. П2 и П3 соответственно.

Полный набор независимых количественных переменных включает следующие показатели:

U_Sh – доля городского населения, в % (имеет картографическое отображение на рис. П4);

NGC_pc – средний за 1883–1892 гг. чистый сбор (за вычетом посева) зерновых культур на душу населения, в пудах;

S – площадь территории уезда, в кв. км;

C_X – широта, в градусах (возрастает к северу);

C_Y – долгота, в градусах (возрастает к востоку);

Центрального статистического комитета Министерства внутренних дел. Санкт-Петербург: Центральный статистический комитет Министерства внутренних дел, 1894. № 34.

²⁰ Указанные возрастные когорты являются наиболее показательными с точки зрения уровня человеческого капитала (соответствуют наименьшей и наибольшей продуктивности трудовой деятельности).

PD – плотность населения (имеет картографическое отображение на рис. П5);

RS_t – время от открытия первой в уезде железнодорожной станции до переписи населения в 1897 г., в количестве лет (развитость железнодорожной инфраструктуры картографически визуализирована на рис. П6);

P_{aa} – осадки в среднем за год, в количестве дней;

T_{aa} – температура в среднем за год, рассчитанная по среднемесячным значениям, в градусах Цельсия.

Приведем полный набор независимых структурных переменных, значения которых выражены в определенных категориях:

river_bank – геоморфологическая характеристика положения относительно водных путей; в качестве контрольной принята категория 3 – отсутствие выраженного характера асимметрии местности;

geo_posit – характер географического положения относительно водных путей; в качестве контрольной принята наиболее массовая категория 1 – речная долина;

gidro – уровень плотности водных потоков и флювиальных форм рельефа; в качестве контрольной категории принят модуль 1 (наименее плотный);

angle – преобладающий угол наклона рельефа; в качестве контрольной принята категория 1 (наиболее ровная поверхность);

soil – характеристика почвенного покрова и ландшафтов (см. рис. П7); в качестве контрольной принята категория 6 – Нечерноземная малоплодородная область.

Анализ включал в себя несколько укрупненных этапов:

1. Определение значимых факторов уровня грамотности и оценивание их роли:

- на малой выборке;
- на средней выборке (полной и с усечением).

2. Определение значимых факторов диспаритета грамотности в разрезе «город/село» и оценивание их роли – на выборке, по которой факторы природной среды дают лучшие результаты.

3. Определение значимых факторов диспаритета грамотности в разрезе «мужчины/женщины» и оценивание их роли – на той же выборке.

Собственно регрессионный анализ на отдельных этапах для каждой выборки проводился в следующем порядке:

1. Первоначально включались только факторы природной среды (в том числе климатические), выраженные количественными переменными, с учетом мультиколлинеарности (с коэффициентами парной корреляции менее 0,7).

2. Факторы природной среды дополнялись экономико-географическими показателями (с учетом мультиколлинеарности), так же выраженными количественными переменными.

3. Проводилось последовательное исключение из модели незначимых количественных переменных ($p < 0,1$). Если на этом этапе в модели не оставалось значимых количественных переменных, характеризующих природную среду, то анализ завершался.

4. Модель, получившаяся в результате предыдущего шага, поочередно дополнялась структурными переменными.

5. Значимые структурные переменные по результатам отбора объединялись с результирующей моделью количественных переменных (шага 4). Если какая-либо переменная оказывалась незначимой, она исключалась, строилась новая регрессия до тех пор, пока все переменные не становились значимыми.

6. Отдельно к результирующей модели с количественными независимыми переменными (шага 4) добавлялись структурные переменные почвенного покрова (данные доступны только для средней выборки с усечением).

Результаты анализа

Основное внимание уделялось построению моделей с зависимой переменной, косвенно характеризующей *уровень* человеческого капитала (укрупненный этап 1).

На *малой выборке* для определения переменных с высоким уровнем взаимосвязи с зависимой переменной логарифма доли грамотных ($LN(L)$) и идентификации возможных проявлений мультиколлинеарности была построена корреляционная матрица количественных переменных (см. табл. 1).

Из табл. 1 видно, что доля городского населения выступает наилучшим прогностическим параметром грамотности. Это также видно из сопоставления рис. П1 и П4. В *малой выборке* колебания логарифма доли городского населения объясняют 74 % колебаний логарифма доли грамотных ($R^2 = 0,86^2 \approx 0,74$). Однако доля городского населения сама является результатом длительных и многофакторных процессов.

Таблица 1

Корреляционная матрица коэффициентов Пирсона между количественными переменными на малой выборке

	$LN(L)$	$LN(U_Sh)$	$LN(NGC_pc)$	$LN(S)$	C_X	C_Y	RS_t	P_aa	T_aa	$LN(PD)$
1	2	3	4	5	6	7	8	9	10	11
$LN(L)$	1									
$LN(U_Sh)$	0,86	1								
$LN(NGC_pc)$	-0,36	-0,31	1							
$LN(S)$	-0,32	-0,41	-0,01	1						
C_X	-0,10	-0,32	-0,06	0,24	1					
C_Y	-0,16	-0,19	0,19	0,48	0,25	1				
RS_t	0,53	0,54	-0,18	-0,42	-0,03	-0,33	1			
P_aa	0,08	-0,03	-0,03	0,04	0,52	-0,08	0,08	1		
T_aa	0,21	0,35	-0,10	-0,31	-0,92	-0,51	0,03	-0,44	1	
$LN(PD)$	0,48	0,59	-0,08	-0,78	-0,33	-0,30	0,64	-0,11	0,31	1

Примечание. Более темным цветом выделены высокие значения тесноты связи между независимыми переменными, менее темным – умеренные.

Климатическая переменная температуры (T_aa) сильно коррелирует с широтой (C_X) и в случае отсутствия данных в бóльших по объему выборках может быть заменена последней. Другая климатическая переменная – осадки за год (P_aa) – с той же широтой (C_X) коррелирует умеренно.

Сильно коррелируют логарифмы площади территории ($LN(S)$) и плотности населения ($LN(PD)$). Использование последней может быть предпочтительнее, поскольку она является частным от двух показателей, оба из которых сравнительно медленно меняются во времени. Кроме того, площадь уезда представляет собой продукт административного распределения властных полномочий, оформленный в виде границ, которые менялись дискретно и на момент переписи могли не отражать их оптимального распределения, в то время как численность населения меняется непрерывно, в направлении к оптимальному размещению.

Коэффициенты корреляции выше 0,5 наблюдаются между логарифмами уровня урбанизации ($LN(U_Sh)$), плотности населения ($LN(PD)$) и

развитостью железнодорожной инфраструктуры (RS_t). В связи с этим предпочтительнее включать их в регрессии попеременно, но не сразу.

На *малой выборке* температура значимо и положительно связана с грамотностью: уровень грамотности выше в более теплых местах. Объем осадков незначим. При этом объясняющая способность такой модели (нормированный $R^2 = 0,0474$) очень низкая.

После исключения незначимых переменных модель, обладающая наилучшим качеством, имеет вид:

$$LN(L) = A + \alpha \cdot RS_t + \beta \cdot LN(NGC_{pc}) + \varepsilon. \quad (1)$$

В табл. 2 приводятся оценки параметров модели (1).

Таблица 2

Регрессионный анализ на малой выборке: зависимость уровня грамотности от физико-географических, экономико-географических количественных факторов

Регрессионная статистика				
Множественный R	0,5951			
R -квадрат	0,3541			
Нормированный R -квадрат	0,3311			
Стандартная ошибка	0,2681			
Наблюдения	59			
Дисперсионный анализ				
	df	F	Значимость F	
Регрессия	2	15,35345	4,83E-06	
Остаток	56			
Итого	58			
	Коэффициенты	Стандартная ошибка	t -статистика	P -значение
Y -пересечение	3,5067	0,1118	31,3708	3,22E-37
RS_t	0,0101	0,0023	4,3780	5,29E-05
$LN(NGC_{pc})$	-0,0916	0,0361	-2,5362	0,014024

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный R -квадрат.

Основной итог анализа на малой выборке (59 уездов) заключается в том, что значимыми оказываются экономико-географические показатели, в то время как физико-географические остаются незначимыми. Также мы видим, что увеличение продолжительности функционирования железнодорожного транспорта связано с повышением уровня грамотности. В то же время, сельскохозяйственная специализация экономики уезда на производстве

зерновых связана с пониженным уровнем грамотности (как было отмечено в литературе на материале западных стран).

На *средней выборке* была построена другая корреляционная матрица количественных переменных (см. табл. 3).

Таблица 3

Корреляционная матрица коэффициентов Пирсона между количественными переменными на средней выборке

	$LN(L)$	$LN(U_Sh)$	$LN(NGC_pc)$	$LN(S)$	C_X	C_Y	RS_t	$LN(PD)$
$LN(L)$	1							
$LN(U_Sh)$	0,47	1						
$LN(NGC_pc)$	-0,16	-0,23	1					
$LN(S)$	-0,08	-0,13	-0,11	1				
C_X	0,31	-0,15	-0,21	0,07	1			
C_Y	-0,35	-0,26	0,09	0,37	0,17	1		
RS_t	0,26	0,48	-0,14	-0,06	0,00	-0,15	1	
$LN(PD)$	-0,04	0,27	0,16	-0,66	-0,41	-0,26	0,26	1

Примечание. Более темным цветом выделены высокие значения тесноты связи между независимыми переменными, менее темным – умеренные.

Как и в *малой выборке*, лучшим предиктором грамотности выступает доля городского населения. Однако в *средней выборке* колебания ее логарифма объясняют только 22 % колебаний логарифма доли грамотных ($R^2 = 0,47^2 \approx 0,22$). Среди независимых эта переменная ($LN(S)$) сильно коррелирует с плотностью населения ($LN(PD)$), однако умеренная корреляция последней с развитостью железнодорожной инфраструктуры (RS_t) становится слабой, что позволяет включать обе переменные в одну регрессию.

На *средней выборке*, как и в случае с *малой*, в первую очередь строилась регрессия с включением только факторов природной среды. В ней географические координаты примерно на 26 % объясняют различия в уровне грамотности между регионами Европейской России. Причем если при движении с запада на восток уровень грамотности имел тенденцию понижаться, то в направлении с юга на север он имел тенденцию возрастать. Это наглядно демонстрирует рис. П1.

После исключения незначимых переменных модель, обладающая наилучшим качеством, принимает следующий вид:

$$LN(L) = A + \alpha \cdot C_X + \beta \cdot C_Y + \gamma \cdot RS_t + \eta \cdot geo_posit_5 + \varepsilon. \quad (2)$$

Оценки ее параметров приведены в табл. 4.

Таблица 4

Регрессионный анализ на средней выборке: набор значимых количественных и структурных факторов уровня грамотности

Регрессионная статистика				
Множественный <i>R</i>	0,6116			
<i>R</i> -квадрат	0,3741			
Нормированный <i>R</i> -квадрат	0,3681			
Стандартная ошибка	0,3579			
Наблюдения	421			
Дисперсионный анализ				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Регрессия	4	62,15997378	3,72E-41	
Остаток	416			
Итого	420			
	Коэффициенты	Стандартная ошибка	<i>t</i> -статистика	<i>P</i> -значение
<i>Y</i> -пересечение	1,1093	0,2762	4,0163	0,0001
<i>C_X</i>	-0,0205	0,0023	-9,0285	0,0000
<i>C_Y</i>	0,0530	0,0051	10,3312	0,0000
<i>RS_t</i>	0,0071	0,0013	5,3077	0,0000
<i>geo_posit = 5</i>	0,7402	0,1108	6,6797	0,0000

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный *R*-квадрат.

Добавление к модели фактора длительности функционирования железнодорожной инфраструктуры повышает ее объясняющую способность с 26 до 30 %. Увеличение данного показателя на 1 год при условно одинаковых географических координатах ведет к повышению натурального логарифма уровня грамотности населения в условном уезде на 0,0071, т.е. уровня грамотности – на 1,01 процентных пунктов (далее – п.п.) ($\approx e^{0,0071} \approx 2,72^{0,0071}$).

Включение хотя бы одной значимой структурной переменной в модель (2) повышает ее объясняющую способность с 30 до 37 %. В данном случае переменная *geo_posit = 5* выражает категорию «морское побережье». Значение соответствующего коэффициента регрессии интерпретируется следующим образом: в уездах на морском побережье, при всех прочих равных, натуральный логарифм уровня грамотности на 0,7402 выше по сравнению с со значением данного показателя в уездах в речных долинах (речная долина принята в качестве базы сравнения как контрольная категория). Соответственно, уровень грамотности населения в уездах на морском побережье выше на 2,10 п.п. ($\approx e^{0,7402} \approx 2,72^{0,7402}$), чем в уездах в речных долинах.

Отдельного рассмотрения заслуживает модель характеристик почвенного покрова (см. рис. П7) как факторов уровня грамотности. С их добавлением к значимым количественным переменным модели (2), ее объясняющая способность радикально поднимается с 30 до 60 %, несмотря на усечение выборки (379 уездов против 421 в полной средней). Результаты регрессионного анализа с участием структурной переменной почвенного покрова, кодированной в комплекс переменных с бинарными значениями 1 и 0, представлены в табл. 5.

Таблица 5

**Результаты регрессионного анализа на средней выборке с усечением:
набор значимых количественных переменных и структурной переменной
характеристики почвенного покрова как факторов уровня грамотности**

Регрессионная статистика				
Множественный R	0,7876			
R -квадрат	0,6203			
Нормированный R -квадрат	0,6046			
Стандартная ошибка	0,2857			
Наблюдения	379			
Дисперсионный анализ				
	df	F	Значимость F	
Регрессия	15	39,53289687	6,31E-67	
Остаток	363			
Итого	378			
	Коэффициенты	Стандартная ошибка	t -статистика	P -значение
Y -пересечение	-0,1449	0,4589	-0,3156	0,7525
CENTROID_X	-0,0299	0,0036	-8,2480	0,0000
CENTROID_Y	0,0830	0,0094	8,8178	0,0000
RS_t	0,0073	0,0012	6,1987	0,0000
Soil = 2	0,5077	0,0900	5,6390	0,0000
Soil = 4	0,8220	0,1458	5,6396	0,0000
Soil = 5	0,2678	0,0756	3,5402	0,0005
Soil = 7	-0,4889	0,1194	-4,0933	0,0001
Soil = 8	-0,3276	0,0657	-4,9876	0,0000
Soil = 9	-0,2309	0,0864	-2,6739	0,0078
Soil = 10	-0,1418	0,0824	-1,7223	0,0859
Soil = 12	-0,8172	0,3141	-2,6018	0,0097
Soil = 13	1,0102	0,2257	4,4750	0,0000

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный R -квадрат.

Как видно из табл. 5, большинство категорий почвенного покрова оказались статистически значимо связаны с уровнем грамотности населения уездов.

В одних значимых категориях уровень грамотности, при всех прочих равных, склонен быть ниже по сравнению с контрольной категорией 6 (Нечерноземная малоплодородная область). Максимальная отрицательная разница отмечена с категорией 12 (Крайний Север, сплошные леса, ограниченное плодородие почвы). Таким образом, в уездах с экстремально низким качеством почвенного покрова препятствия для накопления человеческого капитала оказываются более серьезными. Речь идет об уездах Архангельской губернии, отдельных уездах Вологодской и Олонецкой губерний и одном уезде Пермской губернии.

В других значимых категориях уровень грамотности, при всех прочих равных, имеет тенденцию быть выше по сравнению с контрольной категорией 6.

Укрупненный этап 2 заключался в анализе факторов поселенческого диспаритета «город/село». После пошагового исключения незначимых количественных переменных и отбора значимых кодировок структурных переменных модель принимает следующий вид:

$$LN(SG) = A + \alpha \cdot LN(L) + \beta \cdot C_X + \gamma \cdot C_Y + \eta \cdot LN(NGC_pc) + \mu \cdot RS_t + \varphi \cdot river_bank_2 + \varepsilon. \quad (3)$$

Оценки ее параметров приведены в табл. 6.

В совокупности пять факторов модели объясняют 69 % вариации диспаритета в уровне грамотности между городским и сельским населением уездов Европейской России. Помимо собственно уровня грамотности, значимы обе географические координаты, а также железнодорожная транспортная инфраструктура и отраслевая специализация экономики уезда. Диспаритет «город/село» снижается при движении с запада на восток и возрастает при движении с юга на север. Также данный диспаритет склонен понижаться при усилении аграрно-зерновой специализации уезда.

Добавление лишь одной структурной переменной почти не увеличило объясняющую способность модели (3). Коэффициент при структурной переменной $river_bank = 2$ показывает, что уезды в зоне пойменного берега, при всех прочих равных, склонны иметь диспаритет грамотности ниже примерно на 1,11 пункта ($\approx e^{0,1039} \approx 2,72^{0,1039}$) по сравнению с контрольной категорией (при отсутствии выраженного характера асимметрии местности).

Таблица 6

**Зависимость диспаритета в человеческом капитале (город/село):
набор значимых количественных и структурных переменных**

Регрессионная статистика				
Множественный <i>R</i>	0,8333			
<i>R</i> -квадрат	0,6944			
Нормированный <i>R</i> -квадрат	0,6900			
Стандартная ошибка	0,2095			
Наблюдения	421			
Дисперсионный анализ				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Регрессия	6	156,80685	2,7571E-103	
Остаток	414			
Итого	420			
	Коэффициенты	Стандартная ошибка	<i>t</i> -статистика	<i>P</i> -значение
<i>Y</i> -пересечение	2,4666	0,1886	13,0773	0,0000
<i>LN (L)</i>	-0,7760	0,0273	-28,4455	0,0000
<i>C_X</i>	0,0197	0,0034	5,8601	0,0000
<i>C_Y</i>	-0,0029	0,0015	-2,0087	0,0452
<i>LN (NGC_pc)</i>	-0,0339	0,0202	-1,6792	0,0939
<i>RS_t</i>	0,0044	0,0008	5,4787	0,0000
<i>river_bank = 2</i>	-0,1039	0,0429	-2,4186	0,0160

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный *R*-квадрат.

Как и в случае с анализом факторов уровня грамотности, остановимся на почвенном покрове (см. рис. П7), не все категории которого оказались статистически значимыми. Однако, как и в случае с другими структурными переменными, их включение в модель почти не увеличило ее объясняющую способность. Кроме того, стали незначимыми географические координаты и чистый сбор зерновых на душу населения. Результаты регрессионного анализа с участием характеристик почвенного покрова представлены в табл. 7.

Таблица 7

**Зависимость диспаритета в человеческом капитале (город/село):
набор значимых количественных переменных и структурной переменной
характеристики почвенного покрова**

Регрессионная статистика				
Множественный <i>R</i>	0,8384			
<i>R</i> -квадрат	0,7029			
Нормированный <i>R</i> -квадрат	0,6915			
Стандартная ошибка	0,2075			
Наблюдения	379			
Дисперсионный анализ				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Регрессия	14	61,51979	7,98E-87	
Остаток	364			
Итого	378			
	Коэффициенты	Стандартная ошибка	<i>t</i> -статистика	<i>P</i> -значение
<i>Y</i> -пересечение	3,0852	0,1164	26,5128	0,0000
<i>LN(L)</i>	-0,6847	0,0340	-20,1228	0,0000
<i>RS_t</i>	0,0047	0,0009	5,2931	0,0000
Soil = 1	-0,1299	0,0351	-3,7041	0,0002
Soil = 2	-0,2086	0,0468	-4,4550	0,0000
Soil = 4	-0,2581	0,0839	-3,0773	0,0022
Soil = 5	-0,1090	0,0478	-2,2810	0,0231
Soil = 8	0,0893	0,0460	1,9405	0,0531

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный *R*-квадрат.

На последнем укрупненном этапе 3 проводился анализ факторов гендерного диспаритета (мужчины/женщины). После пошагового исключения незначимых количественных переменных и отбора значимых кодировок структурных переменных модель принимает следующий вид:

$$LN(GG) = A + \alpha \cdot LN(L) + \eta \cdot LN(NGC_{pc}) + \mu \cdot river_bank_1 + \varphi \cdot geo_posit_5 + \omega \cdot gidro_2 + \varepsilon \quad (4)$$

Оценки ее параметров приведены в табл. 8.

Таблица 8

**Зависимость диспаритета в человеческом капитале (мужчины/женщины):
набор значимых количественных и структурных переменных**

Регрессионная статистика				
Множественный <i>R</i>	0,7698			
<i>R</i> -квадрат	0,5926			
Нормированный <i>R</i> -квадрат	0,5877			
Стандартная ошибка	0,3305			
Наблюдения	421			
Дисперсионный анализ				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Регрессия	5	120,7278	1,25271E-78	
Остаток	415			
Итого	420			
	Коэффициенты	Стандартная ошибка	<i>t</i> -статистика	<i>P</i> -значение
<i>Y</i> -пересечение	3,8052	0,1667	22,8284	0,0000
<i>LN (L)</i>	-0,8182	0,0385	-21,2727	0,0000
<i>LN (NGC_pc)</i>	0,0676	0,0315	2,1461	0,0324
<i>river_bank = 1</i>	-0,0988	0,0330	-2,9944	0,0029
<i>geo_posit = 5</i>	-0,2409	0,1068	-2,2562	0,0246
<i>gidro = 2</i>	-0,0750	0,0453	-1,6539	0,0989

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный *R*-квадрат.

Обе географические координаты оказались незначимыми. Добавление трех структурных переменных лишь немного повысило объясняющую способность модели. Отрицательный знак коэффициента при переменной *river_bank = 1* свидетельствует, что гендерный диспаритет, при прочих равных, ниже в зонах коренного берега по сравнению с теми, в которых не выражен характер асимметрии местности. Аналогично, гендерный диспаритет ниже на морском побережье (*geo_posit = 5*) по сравнению с речной долиной; во втором модуле по возрастанию плотности по сравнению с первым (*gidro = 2*).

Как и в случае с анализом факторов уровня грамотности и поселенческого диспаритета, остановимся отдельно на характеристиках почвенного покрова (см. рис. П7), которые оказались статистически значимыми. Результаты регрессионного анализа с участием этой структурной переменной представлены в табл. 9.

Таблица 9

**Зависимость диспаритета в человеческом капитале (мужчины/женщины):
набор значимых количественных переменных и структурной переменной
характеристики почвенного покрова**

Регрессионная статистика				
Множественный <i>R</i>	0,8423			
<i>R</i> -квадрат	0,7095			
Нормированный <i>R</i> -квадрат	0,6983			
Стандартная ошибка	0,2881			
Наблюдения	379			
Дисперсионный анализ				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Регрессия	14	63,50558	1,42E-88	
Остаток	364			
Итого	378			
	Коэффициенты	Стандартная ошибка	<i>t</i> -статистика	<i>P</i> -значение
<i>Y</i> -пересечение	3,9691	0,1809	21,9350	0,0000
<i>LN (S)</i>	-0,8029	0,0449	-17,8840	0,0000
<i>LN (NGC_{pc})</i>	0,0609	0,0333	1,8253	0,0688
Soil = 1	-0,2034	0,0507	-4,0146	0,0001
Soil = 2	-0,3664	0,0674	-5,4393	0,0000
Soil = 3	-0,5816	0,1109	-5,2448	0,0000
Soil = 4	-0,4269	0,1184	-3,6060	0,0004
Soil = 5	-0,5386	0,0656	-8,2113	0,0000
Soil = 7	-0,6249	0,1179	-5,3014	0,0000
Soil = 8	-0,2771	0,0641	-4,3237	0,0000
Soil = 9	-0,2111	0,0727	-2,9020	0,0039
Soil = 10	-0,1438	0,0807	-1,7830	0,0754
Soil = 11	-0,5369	0,1108	-4,8441	0,0000
Soil = 13	-0,6771	0,2078	-3,2577	0,0012

Примечание. Полужирным шрифтом выделен наиболее важный показатель объясняющей способности множественной регрессии – нормированный *R*-квадрат.

Объясняющая способность модели после добавления данной структурной переменной возрастает с 58 до 70 %. Почти все категории почв (кроме 12) значимо отрицательно связаны с гендерным диспаритетом относительно контрольной категории 6.

Таким образом, поселенческий диспаритет в большей степени объясняется количественными переменными (из которых к физико-географическим

относятся широта и долгота), гендерный – структурными (из которых к физико-географическим относятся нахождение в зоне коренного берега, морского побережья, во втором модуле по возрастанию плотности водных потоков и флювиальных неровностей, а также большинство категорий почвенного покрова).

Заключение

С пространственным распределением (по уездам) накопленного в России к 1897 г. уровня человеческого капитала, прокси-индикатором которого в данном случае выступает доля элементарно грамотного населения, статистически значимо ($p < 0,1$) связан ряд существенных факторов природной среды:

- широта (сильно коррелирует с температурой воздуха и умеренно с объемом осадков);
- долгота (умеренно коррелирует с температурой воздуха);
- морское побережье относительно речной долины;
- большинство категорий почвенного покрова.

Ряд существенных факторов природной среды также значимо связан с поселенческим диспаратетом уровня человеческого капитала (в разрезе «город/село»), накопленного к 1897 г.:

- обе координаты (широта и долгота);
- нахождение в зоне пойменного берега (относительно невыраженного характера асимметрии местности);
- некоторые характеристики почвенного покрова.

В меньшей степени факторы природной среды связаны с гендерным диспаратетом человеческого капитала. Из них значимы только структурные переменные, выраженные в следующих категориях:

- нахождение в зоне коренного берега (относительно невыраженного характера асимметрии местности);
- расположение на морском побережье (относительно речной долины);
- уровень плотности водных потоков и флювиальных форм рельефа (второй по возрастанию модуль);
- почвенный покров (в большинстве категорий).

В обоих диспаратетах основным прогностическим параметром выступает общий уровень грамотности. Также значимы факторы, учитывающие технологический уровень и институциональную среду:

- чистый сбор зерновых на душу населения (в большей степени в гендерном);
- уровень развития железнодорожной инфраструктуры (только в поселенческом).

Проведенное исследование подкрепляет количественными аргументами традицию объяснения социально-экономического развития влиянием таких факторов, как климат и другие физико-географические особенности территорий. На российском историческом материале оно показывает, что указанные факторы имели значение не только в условиях преобладания аграрной экономики, но и на ранней стадии индустриализации. В теоретическом аспекте полученные результаты могут служить отправной точкой для определения закономерностей социально-экономического развития на переходных исторических этапах. Приведенные картограммы могут быть использованы в историко-географических и экономико-географических исследованиях отдельных территорий. Построение цифровых карт с тематической нагрузкой позволяет извлечь картометрические показатели изучаемых регионов (площади, зоны контактов, морфологию объектов), а также установить связь с географическими условиями. Представление данных в формате картограммы дает возможность проследить характер распространения определенных показателей в пространстве и влияние соседних регионов.

Разработанные статистические модели зависимости накопления человеческого капитала от характера природной среды описывают эмпирические взаимосвязи, однако не дают представления о динамике этой зависимости и не доказывают наличие причинно-следственных отношений. При этом можно утверждать, что на рассматриваемом историческом этапе технологического развития человеческая деятельность не могла значимо влиять на природно-географическую среду, в то время как люди в той или иной степени могли принимать ее во внимание при выборе мест проживания и видов деятельности. Соответственно, эти модели нуждаются в проверке на данных за иные периоды и по другим локациям. В первую очередь разработанная и апробированная методика анализа может быть реплицирована на исторических данных по регионам Сибири и Дальнего Востока. В качестве другого направления их использования можно предложить расширение круга зависимых переменных, с включением культурных традиций (доля преобладающего вероисповедания), институциональной среды (уровень убийств, концентрация земельной собственности, доля крепостного населения до 1861 г.).

В более подробном рассмотрении нуждаются также механизмы влияния природной среды на формирование человеческого капитала и его пространственную дифференциацию. Заслуживают внимания следующие аспекты этого влияния:

- характер специализации сельскохозяйственного производства;
- характер доступа к водным транспортным путям;

– характер и уровень развития железнодорожной транспортной инфраструктуры, которая в меньшей степени связана с ограничениями со стороны природной среды.

Результаты проведенного анализа подтверждают, что природно-климатические условия по-разному влияли на землеемкие (природоемкие), капиталоемкие, трудоемкие и интеллектуалоемкие виды хозяйственной деятельности.

Приложение

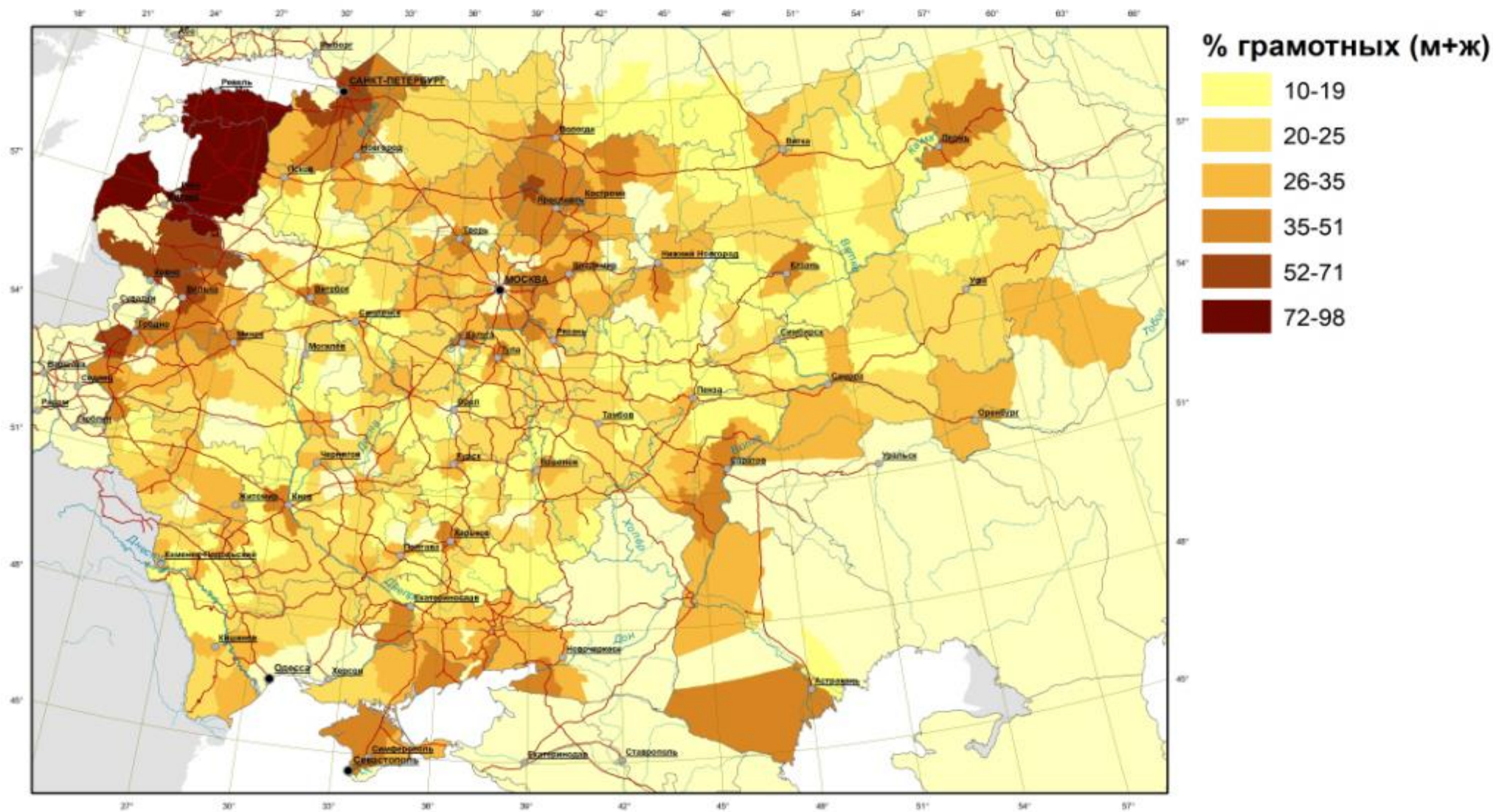


Рис. П1. Уровень грамотности в возрасте 20–29 и 40–49 лет по уездам Европейской России (средняя выборка)

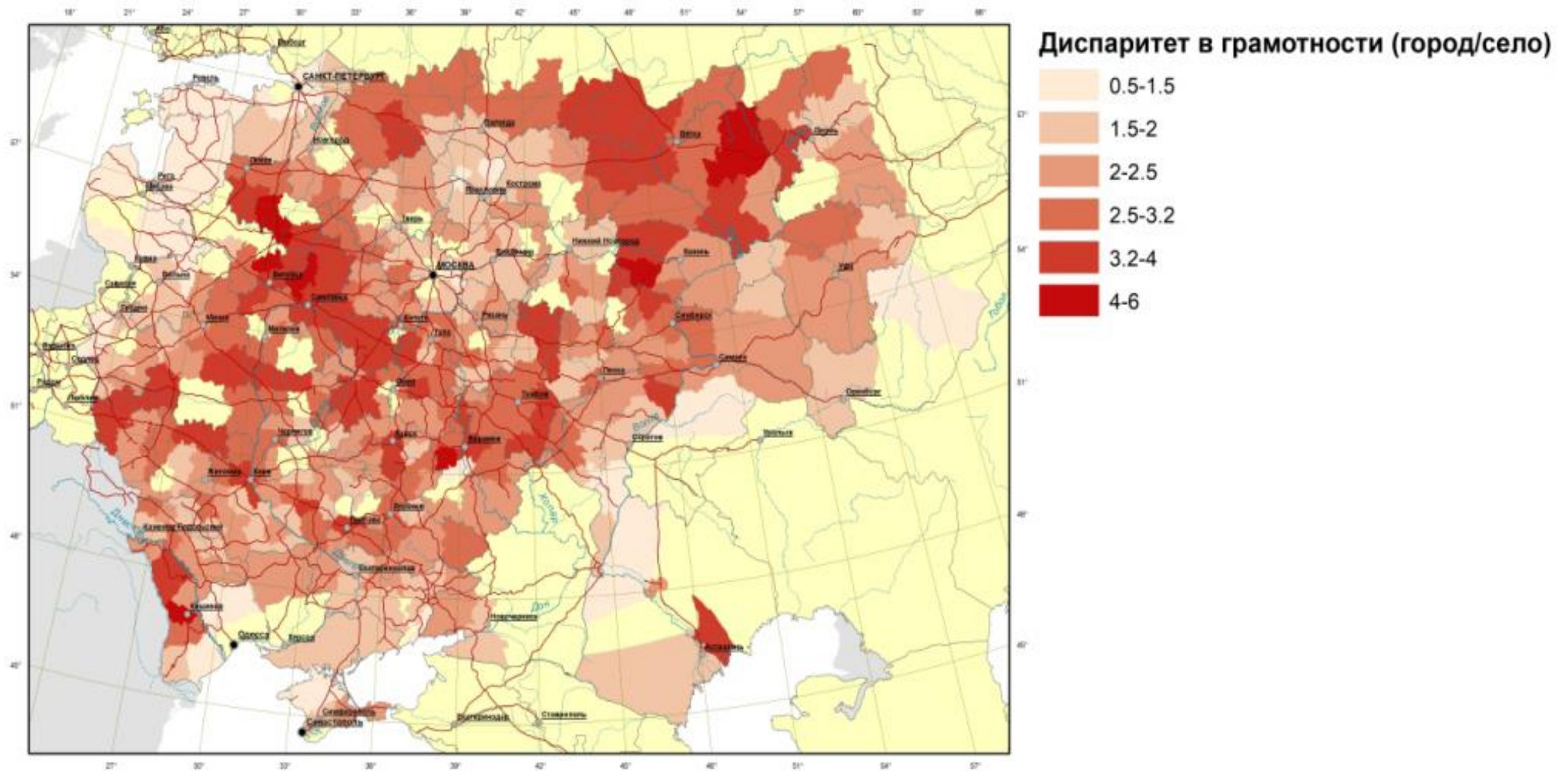


Рис. П2. Поселенческий диспаритет в уровне грамотности возрастных когорт 20–29 и 40–49 лет по уездам Европейской России (средняя выборка)

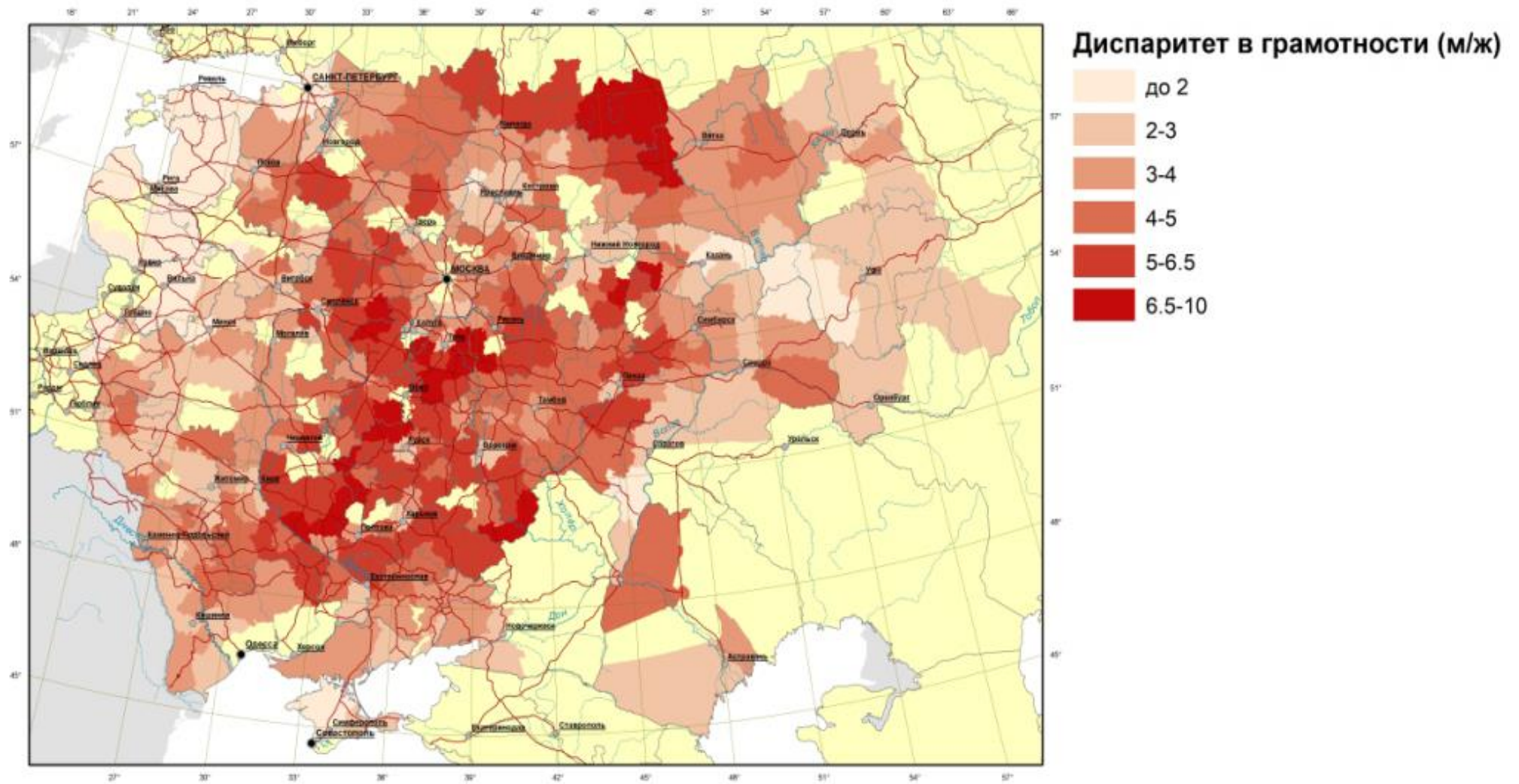


Рис. ПЗ. Гендерный диспаритет в уровне грамотности возрастных когорт 20–29 и 40–49 лет по уездам Европейской России (средняя выборка)

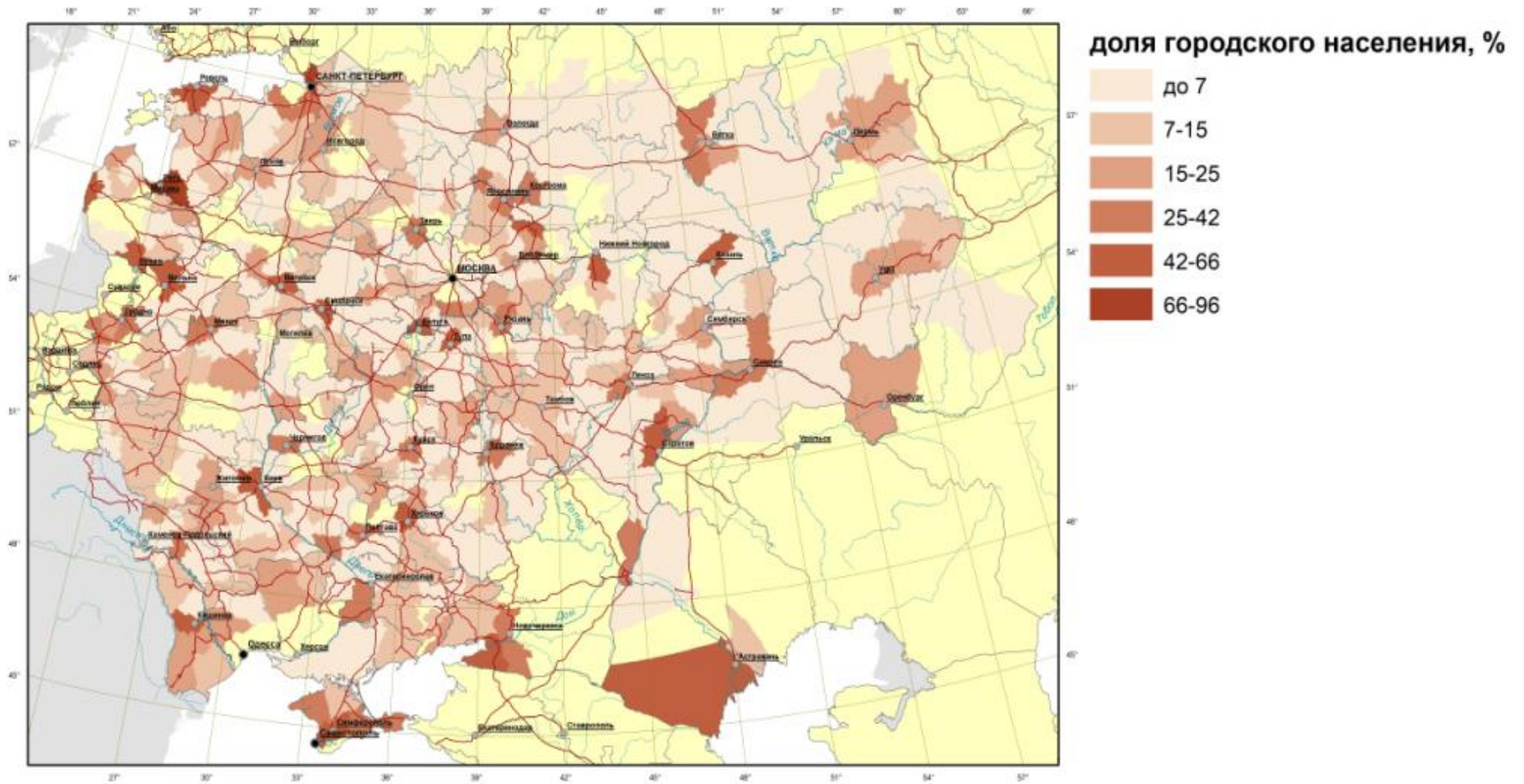


Рис. П4. Уровень урбанизации по уездам Европейской России (средняя выборка)

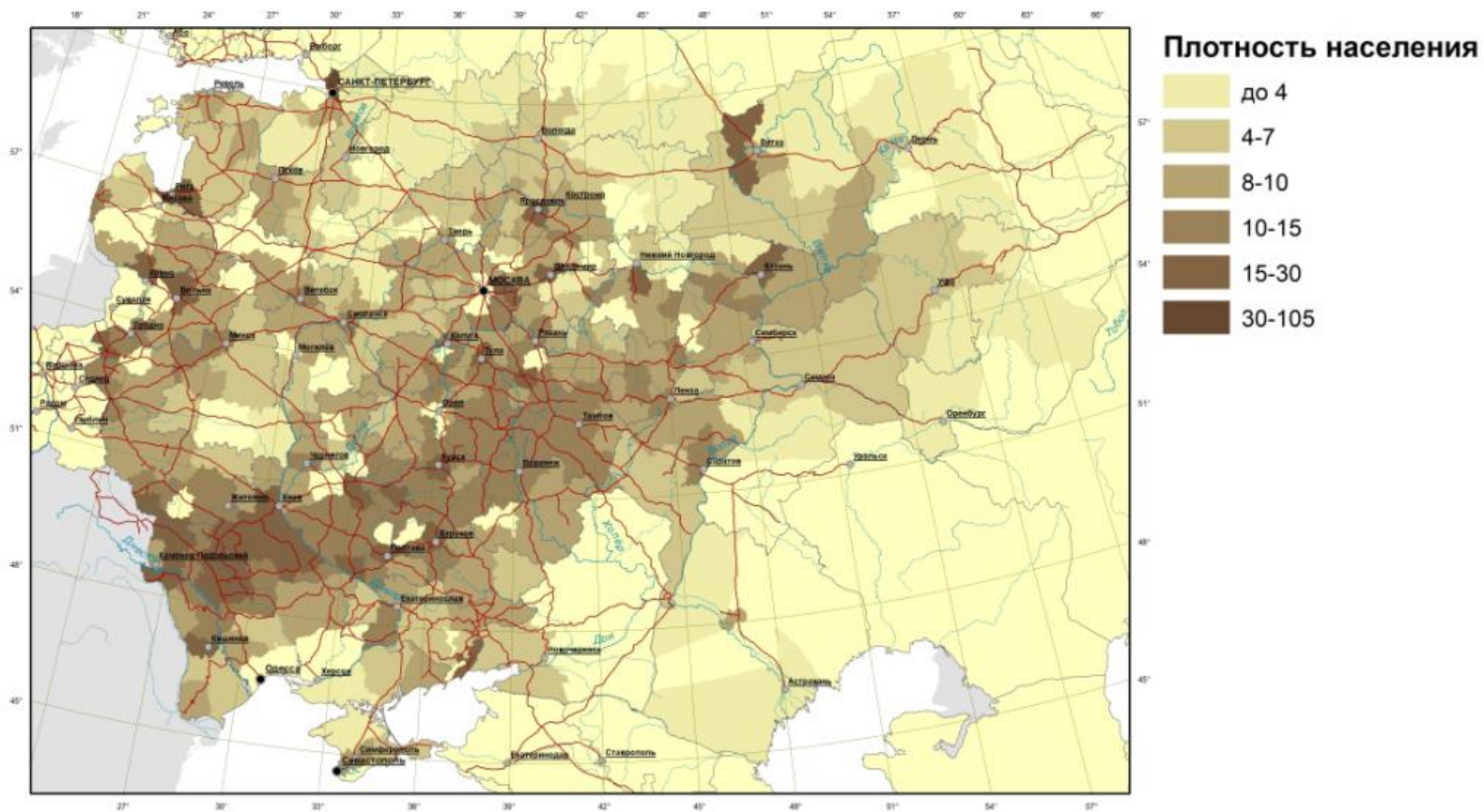


Рис. П5. Плотность населения по уездам Европейской России (средняя выборка)



Рис. Пб. Транспортная инфраструктура, административные границы и центры губерний Европейской России

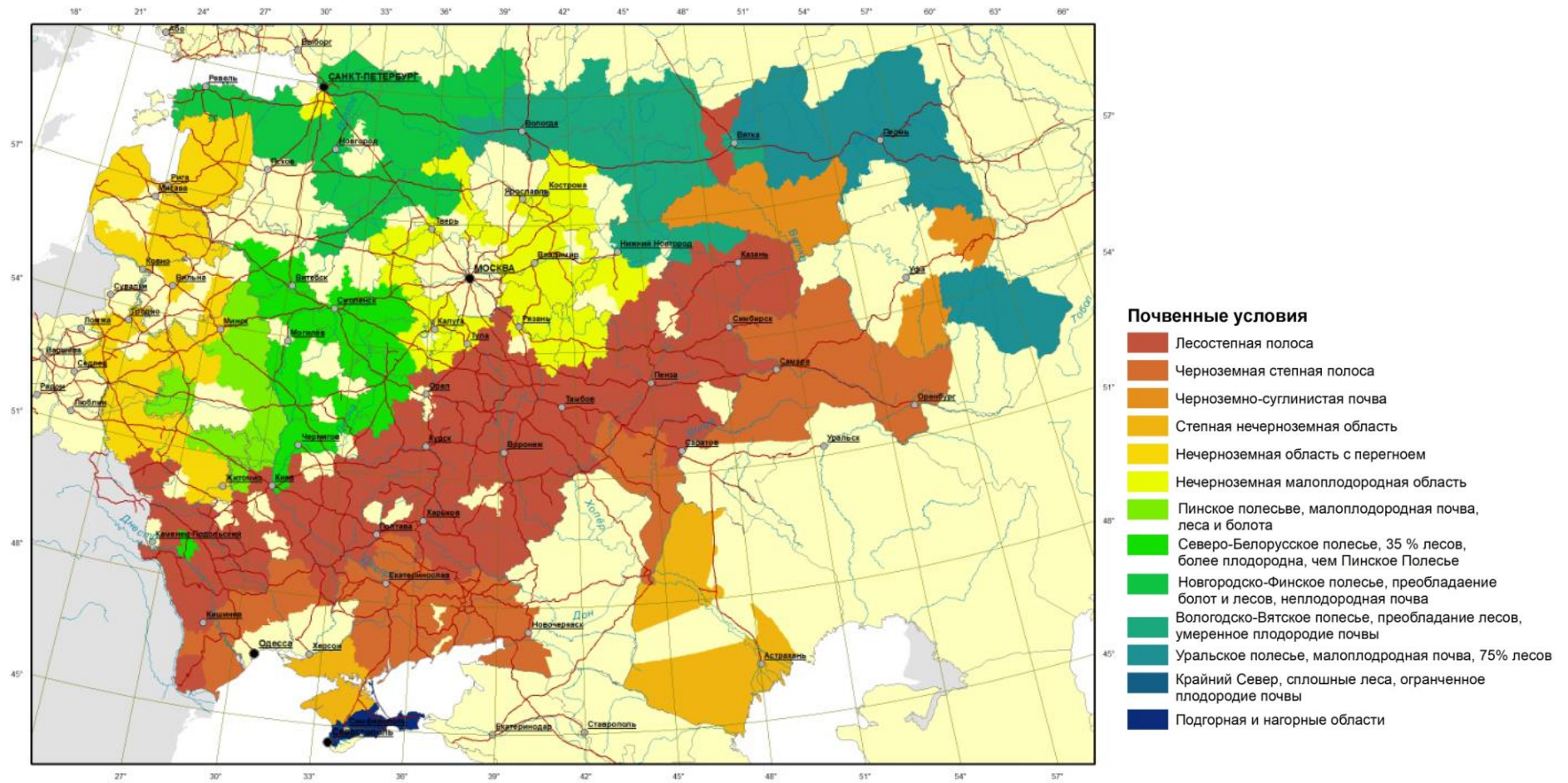


Рис. П7. Почвенный покров по уездам Европейской России (средняя выборка с усечением)



Introduction

The traditions of explaining the processes of socio-economic development by the influence of such factors as climate and other physiographic features of territories are not only widely popular, but they are also confirmed by many empirical studies based mainly on the evidence of foreign countries. In the works on the history of Russia, this issue is interpreted as the influence of natural climatic and geographical factors on the course of national history.

According to R. Pipes,¹ natural conditions contributed to the genesis of the system of institutions of the “patrimonial state” (relations of “property power” in the terminology of post-Marxist institutionalism by R. Nureev and Yu. Latov²). A similar view, largely based on the ideas of V. Klyuchevsky about the decisive role of geographical factors in the Russian historical process under the predominance of the agrarian economy, was presented by L. Milov.³

P. Turchin and S. Nefedov updated the concepts of capacity of an ecological niche and carrying capacity of the environment, which reflect a historical level of interaction between resources and technologies.⁴ In the structural-demographic theory of J. Goldstone,⁵ these concepts were operationalized by such indicators as the population (and separately the elite) in relation to resources and the ratio of the actual population to the upper limit derived from the subsistence minimum combined with available technologies (the level of population pressure on resources).⁶

B. Mironov applied to some empirical evidence that refutes the hypotheses formulated on the basis of geographical determinism, which insists on the decisive role of climate in the history of Russia.⁷

R. Nureev and Yu. Latov largely used quantitative evidence and a number of conceptual ideas of R. Pipes and L. Milov, but the conclusions about the importance of natural-geographical and institutional factors that the researchers reached were closer to the conclusions of B. Mironov. They carried out the idea that the natural environment was a very significant component, considerably influenced the results

¹ R. Pipes, *Russia under the Old Regime* [in Russian], trans. by V. Kozlovskii (Moscow: Nezavisimaya gazeta, 1993), 12–41.

² R.M. Nureev, and Yu.V. Latov, *Economic History of Russia (Studies in Institutional Analysis)* [in Russian] (Moscow: KNORUS, 2016).

³ L.V. Milov, *The Great Russian Ploughman and the Peculiarities of the Russian Historical Process* [in Russian] (Moscow: ROSSPEN, 2001), 4–17, 31–38, 383–85.

⁴ P. Turchin, and S.A. Nefedov, *Secular Cycles* (Princeton, NJ: Princeton University Press, 2009), 6, 18.

⁵ J.A. Goldstone, *Revolution and Rebellion in the Early Modern World* (Berkeley: University of California Press, 1991).

⁶ Turchin, and Nefedov, *Secular Cycles*, 81.

⁷ B.N. Mironov, *The Russian Empire: from Tradition to Modernity* [in Russian], 3 vols. (St Petersburg: DMITRII BULANIN, 2018), vol. 1, 221–50.

of the competition of institutions with historically given alternatives (in the earlier period to a greater extent than in the XX century), but hardly ever determined its outcome.⁸

These works mostly contained a narrative. Quantitative evidence on the problem under consideration without reference to econometric tools was contained to the greatest extent in the works of B. Mironov, P. Turchin and S. Nefedov.

Among the works based on the application of econometric methods, we note the article by J. Baten and R. Hippe.⁹ When testing models of the influence of the natural environment on the inequality of land ownership,¹⁰ they used the materials on European Russia of the period under consideration.

The purpose of our study is to cartographically visualize and statistically test the hypothesis of a significant influence of natural and geographical factors on the formation of human capital of the late Russian Empire, giving their quantitative estimates.

Characteristics of the data from the sources

In the course of the work, the authors analyzed several groups of indicators. Demographic information collected during the First General Census of 1897 was considered as key¹¹: the population and the number of literates in the sections “men / women” and “urban / rural” in the age groups of 20–29 and 40–49 year-olds.¹² Thus, the accumulated human capital is indirectly characterized by the proportion of those who are literate (in this case, those who have the ability to read in their native language).

Another group comprises characteristics of the natural environment. These are physiographic indicators: the area of the territory, longitude, latitude, geomorphological and geographical position regarding waterways, the level of

⁸ Nureev, and Latov, *Economic History of Russia*, 37–46.

⁹ J. Baten, and R. Hippe, “Geography, Land Inequality and Regional Numeracy in Europe in Historical Perspective,” *Journal of Economic Growth*, vol. 23, no. 1 (2018): 79–109.

¹⁰ K.L. Sokoloff, and S.L. Engerman, “History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World,” *Journal of Economic Perspectives*, vol. 14, no. 3 (2000): 217–32; S.L. Engerman, and K.L. Sokoloff, “Factor Endowments, Inequality, and Paths of Development among New World Economies,” *Economía Journal*, vol. 3, no. 1 (2002): 41–109; O. Galor, O. Moav, and D. Vollrath, “Land Inequality and the Emergence of Human Capital Promoting Institutions,” *Review of Economic Studies*, vol. 76, no. 1 (2009): 143–79.

¹¹ N.A. Troinitskii, ed., *The First General Census of the Population of the Russian Empire, 1897* [in Russian], 89 vols. (St Petersburg: Tsentral'nyi statisticheskii komitet Ministerstva vnutrennikh del, 1897–1905). Vols. 1–50, 65, 67, 68, 70.

¹² Provided in spreadsheet format by T. Valetov as the supervisor of A. Degtyareva's bachelor's thesis, 2021.

density of water flows and fluvial landforms,¹³ soil conditions,¹⁴ the prevailing angle of inclination of the relief.¹⁵ They side with climatic indicators: the number of days with precipitation per year¹⁶ and the average annual temperature calculated from monthly data.¹⁷

The geomorphological characteristics of the position regarding the waterways include the following categories:

- 1 – bedrock shore;
- 2 – floodplain shore;
- 3 – the absence of a pronounced character of the asymmetry of the terrain.

The nature of the geographical location (geo_position) of a settlement is identified by the following categories:

- 1 – river valley;
- 2 – estuary, delta;
- 3 – the mouth at the confluence of rivers;
- 4 – lake shore;
- 5 – sea shore.

In the European part of Russia, 13 geographical zones were selected depending on the soil characteristics:

- 1 – Black soil (*chernozemnaya*) non-steppe zone;

¹³ The value characterizes the ratio of the area of the zone to the total number of various forms of water cut and fluvial irregularities. It was calculated on the basis of processing by digital GIS systems the following source: I.A. Strel'bitskii, *Special Map of European Russia (1:420,000). 1865–1871* [in Russian] (Petrograd: Litografiya kartograficheskogo zavedeniya Voenno-topograficheskogo otdela, 1915). Eight modular groups were distinguished as the density increases.

¹⁴ See L.F. de Livron, comp., *Statistical Review of the Russian Empire. Territory* [in Russian] (St Petersburg: Obshchestvennaya pol'za, 1874).

¹⁵ See A.A. Tillo, comp., *Hypsometric map of European Russia: An Experiment in Mapping the Structure of the Surface of European Russia* [in Russian] (St Petersburg: Departament shosseinykh i vodnykh soobshchenii Ministerstva putei soobshcheniya, 1889). It was established on the basis of DEM after converting the original raster into a raster with the values of the largest slopes. The values obtained were arranged in 9 modular groups.

¹⁶ A.A. Tillo, comp., *Atlas of the Distribution of Atmospheric Precipitation in the River Basins of European Russia by Months and for the Whole Year Based on Twenty Years of Observations, 1871–1890* [in Russian] (St Petersburg: Tipo-litografiya K. Birkenfel'da, 1897); S.I. Nebol'sin, *Atlas of maps of the Average Distribution of Atmospheric Precipitation in European Russia. For Individual Months and for a Year, according to Observations of 1888–1912* [in Russian] (Petrograd: Tipografiya “K. Birkenfel'd”, 1916).

¹⁷ See *On the Air Temperature in the Russian Empire, Based on a Critical Examination of Observations, at the Suggestion of the Minister of State Property P.A. Valuev, Developed by the Director of the Main Physical Observatory and a Member of the Imperial Academy of Sciences H. Wild* [in Russian], iss. 2, appendix (St Petersburg: Tipografiya Imperatorskoi Akademii nauk, 1878–1882).

- 2 – Black soil (*chernozemnaya*) steppe zone;
- 3 – Black soil (*chernozemnaya*) loamy area;
- 4 – Steppe non-black soil (*chernozemnaya*) area;
- 5 – Non-black soil (*nechernozemnaya*) area with humus;
- 6 – Non-black soil (*nechernozemnaya*) barren area;
- 7 – Pinsk woodland, barren soil, forests and swamps;
- 8 – North Belarusian woodland, 35% of forests, more fertile than Pinsk woodland;
- 9 – Novgorod-Finnish woodlands, the predominance of swamps and forests, infertile soil;
- 10 – Vologda-Vyatka woodland, forest predominance, moderate soil fertility;
- 11 – Ural woodland, barren soil, 75% of forests;
- 12 – Far North, continuous forests, limited soil fertility;
- 13 – Piedmont and upland areas (without detailed characteristics of soils).

As an indicator of the development of the railway transport infrastructure, the duration of its operation was determined, namely, the time from the opening of the station till the census of 1897.¹⁸

The sectoral specialization of the local economy (namely, the role of grain production) was characterized by the net grain harvest per capita.¹⁹

The county (*uyezd*) was taken as the unit of analysis. The sample of the European part of the country did not include the territories of Finland, Poland, and Transcaucasia, which by 1897 had been part of the Russian Empire for less than 100 years and were the least integrated into the country's socio-economic space.

The “average sample” included 421 uyezds (data are available for most indicators: area, coordinates, grain production per capita and geological conditions, without soils); the “small sample” included 59 uyezds (in addition, data on climate were available). The “average sample with truncation” included 379 uyezds for which, in addition to the full “average sample” indicators, data on soil category were available.

Analysis methods

To achieve the goal, methods of correlation, regression analysis, and cartographic research were used. The first was the premise of the second.

¹⁸ See G.M. Afonina, comp., *Brief Information on the Development of Domestic Railways from 1838 to 1990* [in Russian] (Moscow: Izdatel'stvo Tsentral'noi nauchno-tekhnicheskoi biblioteki zheleznodorozhnogo transporta Rossii, 1995); I.A. Skorino, *Index of Russian Railways* [in Russian] (St Petersburg: Voennaya tipografiya, 1887).

¹⁹ “Average Harvest of Bread and Potatoes for the Decade of 1883–92 in 60 Governorates of European Russia in Relation to National Food” [in Russian], *Chronicle of the Central Statistical Committee of the Ministry of Internal Affairs*, no. 34 (St Petersburg: Tsentral'nyi statisticheskii komitet Ministerstva vnutrennikh del, 1894).

As dependent variables, 1 proxy indicator of the level of human capital was consistently taken (the proportion of literates of both sexes aged 20–29 and 40–49 year-olds²⁰ in a county – *L*) (see the cartographic display in Fig. A1) and 2 indicators characterizing its *intra-territorial inequality* (disparity “urban / rural” – *SG* and “men / women” – *GG*). The disparity in literacy in the context of “urban / rural” was calculated as the ratio of the proportion of literates in urban places to the proportion of literates in rural areas. Gender disparity in literacy was calculated as the ratio of the proportion of literate male population to the proportion of literate female population. Both disparities have cartographic representations in Fig. A2 and A3, respectively.

The full set of independent quantitative variables includes the following indicators:

U_Sh – the share of the urban population, in % (it has a cartographic display in Fig. P4);

NGC_pc – average for 1883–92 net harvest (minus sowing) of grain crops per capita, in poods;

S – the area of the uyezd, in sq. km;

C_X – latitude, in degrees (increasing to the north);

C_Y – longitude, in degrees (increasing towards the east);

PD – population density (has a cartographic display in Fig. A5);

RS_t – time from the opening of the first railway station in the uyezd until the population census in 1897, in the number of years (the development of the railway infrastructure is cartographically visualized in Fig. A6);

P_aa – annual average precipitation, in number of days;

T_aa – average temperature for the year, calculated from average monthly values, in degrees Celsius.

Here is a complete set of independent indicator (dummy) variables, the values of which are expressed in certain categories:

river_bank – geomorphological characteristic of the position regarding the waterways; category 3, the absence of a pronounced character of the asymmetry of the terrain, was adopted as a control category;

geo_posit – the nature of the geographical position regarding the waterways; the most massive category 1, the river valley, was taken as a control category;

gidro – density level of water flows and fluvial landforms; module 1 (least dense) was adopted as a control category;

angle – the prevailing angle of inclination of the relief; category 1 (the most even surface) was adopted as a control category;

²⁰ The indicated age groups are the most indicative in terms of the level of human capital (they correspond to the lowest and highest productivity of labor).

soil – characteristics of the soil cover and landscapes (see Fig. A7); category 6 – Non-black soil (*nechernozemnaya*) barren area was adopted as a control category.

The analysis included several enlarged stages:

1. Identification of significant factors of literacy level and assessment of their role:

- on a small sample;
- on the average sample (full and truncated).

2. Determination of significant factors of literacy disparity in the context of “urban / rural” and assessment of their role on the sample, according to which environmental factors give the best results.

3. Determination of significant factors of literacy disparity in the context of “men / women” and assessment of their role on the same sample.

The actual regression analysis at separate stages for each sample was carried out in the following order:

1. Initially, only environmental factors (including climatic ones) expressed by quantitative variables were included, taking into account multicollinearity (with pair correlation coefficients less than 0.7).

2. Environmental factors were supplemented by economic and geographical indicators (taking into account multicollinearity), also expressed by quantitative variables.

3. A consistent exclusion of insignificant quantitative variables from the model ($p < 0.1$) was carried out. If at this stage there were no significant quantitative variables characterizing the natural environment in the model, the analysis was completed.

4. The model resulting from the previous step was alternately supplemented with dummy variables.

5. Significant dummy variables according to the selection results were combined with the resulting model of quantitative variables (step 4). If any variable turned out to be insignificant, it was excluded, and a new regression was built until all the variables became significant.

6. Separately, dummy variables of the soil cover were added to the resulting model with quantitative independent variables (step 4) (data are available only for the average sample with truncation).

Analysis results

The main attention was paid to the construction of models with a dependent variable that indirectly characterizes the level of human capital (enlarged stage 1).

A correlation matrix of quantitative variables was constructed *on a small sample* to determine variables with a high level of correlation with the dependent variable of the logarithm of the proportion of literate ($LN(L)$) and identify possible manifestations of multicollinearity (see Table 1).

Table 1

**Correlation matrix of Pearson coefficients between quantitative variables
on a small sample**

	$LN(L)$	$LN(U_{Sh})$	$LN(NGC_{pc})$	$LN(S)$	C_X	C_Y	RS_t	P_{aa}	T_{aa}	$LN(PD)$
1	2	3	4	5	6	7	8	9	10	11
$LN(L)$	1									
$LN(U_{Sh})$	0.86	1								
$LN(NGC_{pc})$	-0.36	-0,31	1							
$LN(S)$	-0.32	-0.41	-0.01	1						
C_X	-0.10	-0.32	-0.06	0.24	1					
C_Y	-0.16	-0.19	0.19	0.48	0.25	1				
RS_t	0.53	0.54	-0.18	-0.42	-0.03	-0.33	1			
P_{aa}	0.08	-0.03	-0.03	0.04	0.52	-0.08	0.08	1		
T_{aa}	0.21	0.35	-0.10	-0.31	-0.92	-0.51	0.03	-0.44	1	
$LN(PD)$	0.48	0.59	-0.08	-0.78	-0.33	-0.30	0.64	-0.11	0.31	1

Note. The darker color highlights the high values of the closeness of the relationship between the independent variables; the less dark, the moderate ones.

From the Table 1 it can be seen that the share of the urban population is the best predictive parameter of literacy. This is also evident from the comparison of Fig. A1 and A4. On a small sample, fluctuations in the logarithm of the urban population share explain 74% of the fluctuations in the logarithm of the literate share ($R^2 = 0.86^2 \approx 0.74$). However, the share of the urban population itself is the result of long and multifactorial processes.

The climatic temperature variable (T_{aa}) strongly correlates with latitude (C_X) and in the absence of data in larger samples, it can be replaced by the latter. Another climatic variable – precipitation per year (P_{aa}) – correlates moderately with the same latitude (C_X).

Logarithms of area ($LN(S)$) and population density ($LN(PD)$) strongly correlate. The use of the latter may be preferable, since it is a quotient of the two indicators, both of which change relatively slowly over time. In addition, the area of an uyezd is a product of the administrative distribution of powers, designed in the form of boundaries that changed discretely and at the time of the census could not reflect

their optimal distribution, while the population is changing continuously in the direction of optimal placement.

Correlation coefficients above 0.5 are observed between the logarithms of the level of urbanization ($LN(U_Sh)$), population density ($LN(PD)$) and the development of railway infrastructure (RS_t). In this regard, it is preferable to include them in the regressions alternately, but not simultaneously.

On a *small sample*, temperature is significantly and positively associated with literacy: the literacy rate is higher in warmer places. The amount of precipitation is insignificant. At the same time, the explanatory power of such a model (normalized $R^2 = 0.0474$) is very low.

After excluding insignificant variables, the model with the best quality has the form:

$$LN(L) = A + \alpha \cdot RS_t + \beta \cdot LN(NGC_pc) + \varepsilon. \quad (1)$$

Table 2 provides estimates of the parameters of the model (1).

Table 2

Regression analysis on a small sample: the dependence of the literacy level on physiographic and economic-geographical quantitative factors

Regression statistics				
Multiple R	0.5951			
R-square	0.3541			
Normalized R-square	0.3311			
Standard error	0.2681			
Observations	59			
Analysis of variance				
	<i>df</i>	<i>F</i>	Significance <i>F</i>	
Regression	2	15.35345	4.83E-06	
Residual	56			
Total	58			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
Y-intercept	3.5067	0.1118	31.3708	3.22E-37
RS_t	0.0101	0.0023	4.3780	5.29E-05
$LN(NGC_pc)$	-0.0916	0.0361	-2.5362	0.014024

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

The main result of the analysis on a small sample (59 uyezds) is that economic and geographical indicators turn out to be significant, while physiographic indicators remain insignificant. We also see that an increase in the duration of the operation of railway transport is associated with an increase in the level of literacy. At the same time, the agricultural specialization of the county's economy in grain production is

associated with a reduced level of literacy (as noted in the literature based on the material of Western countries).

Another correlation matrix of quantitative variables was constructed on the average sample (see Table 3).

Table 3

**Correlation matrix of Pearson coefficients between quantitative variables
on the average sample**

	<i>LN (L)</i>	<i>LN (U_Sh)</i>	<i>LN (NGC_pc)</i>	<i>LN (S)</i>	<i>C_X</i>	<i>C_Y</i>	<i>RS_t</i>	<i>LN (PD)</i>
<i>LN (L)</i>	1							
<i>LN (U_Sh)</i>	0.47	1						
<i>LN (NGC_pc)</i>	-0.16	-0.23	1					
<i>LN (S)</i>	-0.08	-0.13	-0.11	1				
<i>C_X</i>	0.31	-0.15	-0.21	0.07	1			
<i>C_Y</i>	-0.35	-0.26	0.09	0.37	0.17	1		
<i>RS_t</i>	0.26	0.48	-0.14	-0.06	0.00	-0.15	1	
<i>LN (PD)</i>	-0.04	0.27	0.16	-0.66	-0.41	-0.26	0.26	1

Note. The darker color highlights the high values of the closeness of the relationship between the independent variables; the less dark, the moderate ones.

As with a *small sample*, the best predictor of literacy is share of the urban population. However, on the average sample, fluctuations in its logarithms explain only 22% of fluctuations in the logarithm of the literacy rate ($R^2 = 0.472 \approx 0.22$). Among the independents, this variable (*LN (S)*) is highly correlated with population density (*LN (PD)*), but the latter's moderate correlation with railway infrastructure development (*RS_t*) becomes weak, which allows both variables to be included in one regression.

On the average sample, as in the case of a small one, regression was primarily built with the inclusion of only natural environment factors. In it, geographical coordinates account for about 26% of the differences in the level of literacy between the regions of European Russia. Moreover, while the literacy rate tended to decrease when moving from west to east, it tended to increase in the direction from south to north. This is clearly demonstrated by Fig. A1.

After excluding insignificant variables, the model with the best quality takes the following form:

$$LN(L) = A + \alpha \cdot C_X + \beta \cdot C_Y + \gamma \cdot RS_t + \eta \cdot geo_posit_5 + \varepsilon. \quad (2)$$

Estimates of its parameters are given in Table 4.

Table 4

Regression analysis on the average sample: a set of significant quantitative and qualitative factors of literacy level

Regression statistics				
Multiple <i>R</i>	0.6116			
<i>R</i> -square	0.3741			
Normalized <i>R</i> -square	0.3681			
Standard error	0.3579			
Observations	421			
Analysis of variance				
	<i>df</i>	<i>F</i>	Significance <i>F</i>	
Regression	4	62.15997378	3.72E-41	
Residual	416			
Total	420			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
<i>Y</i> -intercept	1.1093	0.2762	4,0163	0,0001
<i>C_X</i>	-0.0205	0.0023	-9,0285	0,0000
<i>C_Y</i>	0.0530	0.0051	10,3312	0,0000
<i>RS_t</i>	0.0071	0.0013	5,3077	0,0000
<i>geo_posit = 5</i>	0.7402	0.1108	6,6797	0,0000

Note. The most important indicator of the explanatory power of multiple regression (the normalized *R*-square) is in bold.

Adding the duration factor of the railway infrastructure to the model increases its explanatory power from 26 to 30%. An increase in this indicator for 1 year with conditionally identical geographical coordinates leads to an increase in the natural logarithm of the literacy rate of the population in the conditional uyezd by 0.0071, i.e. the increase of literacy rate by 1.01 percentage points (hereinafter referred to as pp) ($\approx e^{0,0071} \approx 2,72^{0,0071}$).

The inclusion of at least one significant dummy variable in the model (2) increases its explanatory power from 30 to 37%. In this case, the variable *geo_posit = 5* expresses the category “sea shore.” The value of the corresponding regression coefficient is interpreted as follows: in uyezds on the seashore, with all other things being equal, the natural logarithm of the literacy rate tends to be 0.7402 higher compared to the value of this indicator in counties in river valleys (the river valley is accepted as a comparison base as a control category). Accordingly, the literacy rate of the population in the counties on the seashore is 2.10 pp ($\approx e^{0,7402} \approx 2,72^{0,7402}$) higher than in the uyezds in the river valleys.

The model of soil cover characteristics deserves separate consideration (see Fig. A7) as factors of literacy level. With their addition to the significant quantitative

variables of the model (2), its explanatory power radically rises from 30 to 60%, despite the truncation of the sample (379 uyezds versus 421 in the full average). The results of regression analysis involving a dummy variable of soil cover encoded into a complex of variables with binary values of 1 and 0 are presented in Table 5.

Table 5

Results of regression analysis on the average sample with truncation: a set of significant quantitative variables and a dummy variable of soil cover characteristics as factors of literacy level

Regression statistics				
Multiple R	0.7876			
R-square	0.6203			
Normalized R-square	0.6046			
Standard error	0.2857			
Observations	379			
Analysis of variance				
	df	F	Significance F	
Regression	15	39.53289687	6.31E-67	
Residual	363			
Total	378			
	Coefficients	Standard error	t-statistics	P-value
Y- intercept	-0.1449	0.4589	-0.3156	0.7525
CENTROID_X	-0.0299	0.0036	-8.2480	0.0000
CENTROID_Y	0.0830	0.0094	8.8178	0.0000
RS_t	0.0073	0.0012	6.1987	0.0000
Soil = 2	0.5077	0.0900	5.6390	0.0000
Soil = 4	0.8220	0.1458	5.6396	0.0000
Soil = 5	0.2678	0.0756	3.5402	0.0005
Soil = 7	-0.4889	0.1194	-4.0933	0.0001
Soil = 8	-0.3276	0.0657	-4.9876	0.0000
Soil = 9	-0.2309	0.0864	-2.6739	0.0078
Soil = 10	-0.1418	0.0824	-1.7223	0.0859
Soil = 12	-0.8172	0.3141	-2.6018	0.0097
Soil = 13	1.0102	0.2257	4.4750	0.0000

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

As can be seen from Table 5, most of the categories of soil cover were statistically significantly associated with the level of literacy of the uyezds' population.

In some significant categories, the literacy rate, with all other things being equal, tends to be lower compared to the control category 6 (Non-black soil (*nechernozemnaya*) barren area). The maximum negative difference was noted with

category 12 (the Far North, solid forests, limited soil fertility). Thus, in the uyezds with extremely low quality of soil cover, the obstacles to the accumulation of human capital are more serious. We are talking about the uyezds of Arkhangelsk Governorate, some uyezds of Vologda Governorate and Olonets Governorate and one uyezd of Perm Governorate.

In other significant categories, the literacy rate, with all other things being equal, tends to be higher compared to the control category 6.

The enlarged stage 2 consisted in the analysis of factors of settlement *disparity* “urban / rural”. After step-by-step elimination of insignificant quantitative variables and selection of significant encodings of dummy variables, the model takes the following form:

$$LN(SG) = A + \alpha \cdot LN(L) + \beta \cdot C_X + \gamma \cdot C_Y + \eta \cdot LN(NGC_pc) + \mu \cdot RS_t + \varphi \cdot river_bank_2 + \varepsilon. \quad (3)$$

Estimates of its parameters are given in Table 6.

Table 6

Human capital disparity relationship (urban/rural): a set of significant quantitative and dummy variables

Regression statistics				
Multiple R	0.8333			
R-square	0.6944			
Normalized R-square	0.6900			
Standard error	0.2095			
Observations	0.8333			
Analysis of variance				
	<i>df</i>	<i>F</i>	Significance <i>F</i>	
Regression	6	156.80685	2.7571E-103	
Residual	414			
Total	420			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
<i>Y</i> -intercept	2.4666	0.1886	13.0773	0.0000
<i>LN(L)</i>	-0.7760	0.0273	-28.4455	0.0000
<i>C_X</i>	0.0197	0.0034	5.8601	0.0000
<i>C_Y</i>	-0.0029	0.0015	-2.0087	0.0452
<i>LN(NGC_pc)</i>	-0.0339	0.0202	-1.6792	0.0939
<i>RS_t</i>	0.0044	0.0008	5.4787	0.0000
<i>river_bank = 2</i>	-0.1039	0.0429	-2.4186	0.0160

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

Together, the five factors of the model explain 69% of the variation in the disparity in literacy rates between the urban and rural populations of the counties of European Russia. In addition to the level of literacy itself, both geographic coordinates are of significance as well as the railway transport infrastructure and the sectoral specialization of the county's economy. The urban / rural disparity decreases when moving from west to east and increases when moving from south to north. Also, this disparity tends to decrease with the strengthening of the agricultural and grain specialization of the county.

The addition of only one dummy variable did little to increase the explanatory power of model (3). The coefficient on the dummy variable *river_bank* = 2 shows that counties in the floodplain zone, all other things being equal, tend to have a literacy disparity lower by about 1.11 points ($\approx e^{0.1039} \approx 2.72^{0.1039}$) compared to the control category (in the absence of the pronounced nature of the asymmetry of the terrain).

As in the case of the analysis of literacy level factors, let us dwell on the soil cover (see Fig. A7). Not all of its categories turned out to be statistically significant. However, as with other dummy variables, their inclusion in the model did little to increase its explanatory power. In addition, geographic coordinates and net grain harvest per capita became insignificant. The results of the regression analysis involving the characteristics of the soil cover are presented in Table 7.

Table 7

**Dependence of disparity in human capital (urban / rural):
set of significant quantitative variables and dummy variable of soil cover characteristic**

Regression statistics				
Множественный R	0.8384			
Multiple R	0.7029			
R-square	0.6915			
Standard error	0.2075			
Observations	379			
Analysis of variance				
	<i>df</i>	<i>F</i>	Значимость <i>F</i>	
Regression	14	61,51979	7,98E-87	
Residual	364			
Total	378			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
<i>Y</i> -intercept	3.0852	0.1164	26.5128	0.0000
<i>LN</i> (<i>L</i>)	-0.6847	0.0340	-20.1228	0.000
<i>RS_t</i>	0.0047	0.0009	5.2931	0.0000
Soil = 1	-0.1299	0.0351	-3.7041	0.0002

	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
Soil = 2	-0.2086	0.0468	-4.4550	0.0000
Soil = 4	-0.2581	0.0839	-3.0773	0.0022
Soil = 5	-0.1090	0.0478	-2.2810	0.0231
Soil = 8	0.0893	0.0460	1.9405	0.0531

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

At the last enlarged stage 3, an analysis of gender *disparity factors (men / women)* was carried out. After stepwise exclusion of insignificant quantitative variables and selection of significant encodings of dummy variables, the model takes the following form:

$$LN(GG) = A + \alpha \cdot LN(L) + \eta \cdot LN(NGC_{pc}) + \mu \cdot river_bank_1 + \varphi \cdot geo_posit_5 + \omega \cdot gidro_2 + \varepsilon. \quad (4)$$

Estimates of its parameters are given in Table 8.

Table 8

Disparity relationship in human capital (men / women): a set of significant quantitative and dummy variables

Regression statistics				
Multiple R	0.7698			
R-square	0.5926			
Normalized R-square	0.5877			
Standard error	0.3305			
Observations	421			
Analysis of variance				
	<i>df</i>	<i>F</i>	Significance <i>F</i>	
Regression	5	120.7278	1.25271E-78	
Residual	415			
Total	420			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
<i>Y</i> - intercept	3.8052	0.1667	22.8284	0.0000
<i>LN(L)</i>	-0.8182	0.0385	-21.2727	0.0000
<i>LN(NGC_{pc})</i>	0.0676	0.0315	2.1461	0.0324
<i>river_bank = 1</i>	-0.0988	0.0330	-2.9944	0.0029
<i>geo_posit = 5</i>	-0.2409	0.1068	-2.2562	0.0246
<i>gidro = 2</i>	-0.0750	0.0453	-1.6539	0.0989

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

Both geographic coordinates turned out to be insignificant. The addition of three dummy variables only slightly increased the explanatory power of the model. The

negative sign of the coefficient for the variable *river_bank* = 1 indicates that gender disparity, other things being equal, is lower in the bedrock areas compared to those in which the character of the terrain asymmetry is not expressed. Similarly, gender disparity is lower on the seashore (*geo_posit* = 5) compared to the river valley; in the second module in increasing density compared to the first (*gidro* = 2).

As in the case with the analysis of the factors of literacy and settlement disparity, we will dwell separately on the characteristics of the soil cover (see Fig. A7), which turned out to be statistically significant. The results of the regression analysis involving this dummy variable are presented in Table 9.

Table 9

Dependence of disparity in human capital (men / women): a set of significant quantitative variables and dummy variable characteristics of the soil cover

Regression statistics				
Multiple R	0.8423			
R-square	0.7095			
Normalized R-square	0.6983			
Standard error	0.2881			
Observations	379			
Analysis of variance				
	<i>df</i>	<i>F</i>	Significance <i>F</i>	
Regression	14	63.50558	1.42E-88	
Residual	364			
Total	378			
	Coefficients	Standard error	<i>t</i> -statistics	<i>P</i> -value
<i>Y</i> -intercept	3.9691	0.1809	21.9350	0.0000
<i>LN</i> (<i>S</i>)	-0.8029	0.0449	-17.8840	0.0000
<i>LN</i> (<i>NGC_pc</i>)	0.0609	0.0333	1.8253	0.0688
Soil = 1	-0.2034	0.0507	-4.0146	0.0001
Soil = 2	-0.3664	0.0674	-5.4393	0.0000
Soil = 3	-0.5816	0.1109	-5.2448	0.0000
Soil = 4	-0.4269	0.1184	-3.6060	0.0004
Soil = 5	-0.5386	0.0656	-8.2113	0.0000
Soil = 7	-0.6249	0.1179	-5.3014	0.0000
Soil = 8	-0.2771	0.0641	-4.3237	0.0000
Soil = 9	-0.2111	0.0727	-2.9020	0.0039
Soil = 10	-0.1438	0.0807	-1.7830	0.0754
Soil = 11	-0.5369	0.1108	-4.8441	0.0000
Soil = 13	-0.6771	0.2078	-3.2577	0.0012

Note. The most important indicator of the explanatory power of multiple regression (the normalized R-square) is in bold.

The explanatory power of the model after adding this dummy variable increases from 58 to 70%. Almost all soil categories (except 12) are significantly negatively associated with gender disparity relative to the control category 6.

Thus, the settlement disparity is largely explained by quantitative variables (of which latitude and longitude belong to the physiographic ones); gender disparity, by dummy variables (of which the physiographic ones include staying in the area of bedrock shore, the seashore, the second module according to the increasing density of water flows and fluvial irregularities, as well as most categories of soil cover).

Conclusion

The spatial distribution (by uyezds) of the level of human capital accumulated in Russia by 1897, the proxy indicator of which in this case is the proportion of the elementary literate population, is statistically significantly ($p < 0.1$) associated with a number of essential factors of the natural environment:

- latitude (correlates strongly with air temperature and moderately with precipitation);
- longitude (moderately correlates with air temperature);
- the sea shore relative to the river valley;
- most categories of soil cover.

A number of essential factors of natural environment are also significantly associated with the settlement disparity in the level of human capital (in the context of “city / village”) accumulated by 1897:

- both coordinates (latitude and longitude);
- staying in the area of the floodplain (relative to unexpressed nature of the asymmetry of the terrain);
- some characteristics of the soil cover.

To a lesser extent, factors of natural environment are associated with the gender disparity of human capital. Of these, only structural variables expressed in the following categories are significant:

1. staying in the area of the bedrock coast (relative to unexpressed nature of the asymmetry of the terrain);
2. location on the seashore (relative to the river valley);
3. the level of density of water flows and fluvial landforms (the second ascending module);
4. soil cover (in most categories).

In both disparities, the main predictive parameter is the overall level of literacy. Factors that take into account the technological level and institutional environment are also significant:

- net grain harvest per capita (mostly in the gender disparity regressions);

– the level of development of the railway infrastructure (only in the settlement disparity regressions).

The conducted research supports the tradition of explaining socio-economic development by the influence of factors such as climatic and other physiographic features of territories with quantitative arguments. Based on Russian historical evidence, it shows that these factors were important not only in the conditions of the predominance of an agrarian economy but also at an early stage of industrialization. In a theoretical aspect, the results obtained can serve as a starting point for determining the patterns of socio-economic development at transitional historical stages. The given cartograms can be used in historical-geographical and economic-geographical researches of particular territories. The construction of topical digital maps allows extracting cartometric indicators of the regions under study (area, contact zones, morphology of objects) as well as establishing a connection with geographical conditions. The presentation of data in the cartogram format makes it possible to trace the nature of the spatial distribution of certain indicators and the influence of neighboring regions.

The developed statistical models of the dependence of human capital accumulation on the natural environment describe empirical relationships, but do not give an idea of the dynamics and do not prove the existence of causal relationships. At the same time, it can be argued that at the historical stage of technological development under consideration, human activity could not significantly affect the natural and geographical environment, while people could take it into account to one degree or another when choosing places of residence and types of activities. Accordingly, these models need to be tested on data for other periods and for other locations. First of all, the developed and tested method of analysis can be replicated on historical data for the regions of Siberia and the Far East. As another line of their use, we can suggest expanding the range of dependent variables, including cultural traditions (the share of the predominant religion), the institutional environment (the murder rate, the concentration of land ownership, the share of the serf population before 1861).

The mechanisms of the influence of natural environment on the formation of human capital and its spatial differentiation also need to be considered in more detail. The following aspects of this influence deserve attention:

1. the pattern of specialization of agricultural production;
2. the pattern of access to waterways;
3. the pattern and level of development of the railway transport infrastructure, which is less associated with environmental constraints.

The results of the analysis confirm that natural and climatic conditions had different effects on land-intensive (nature-intensive), capital-intensive, labor-intensive and knowledge-intensive types of economic activity.

Appendix

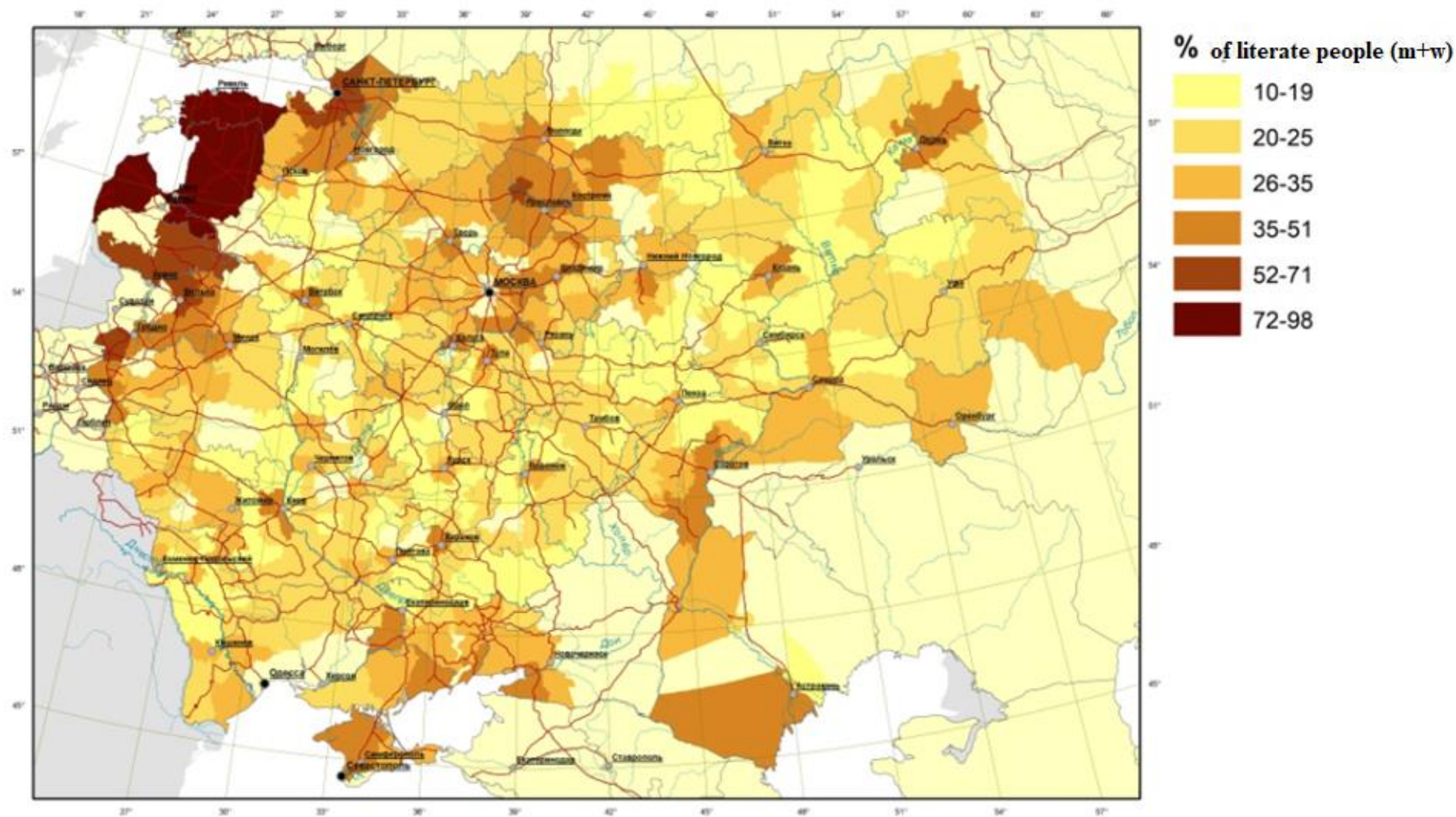


Fig. A1. Literacy rate at the age of 20–29 and 40–49 years by uyezds of European Russia (average sample)

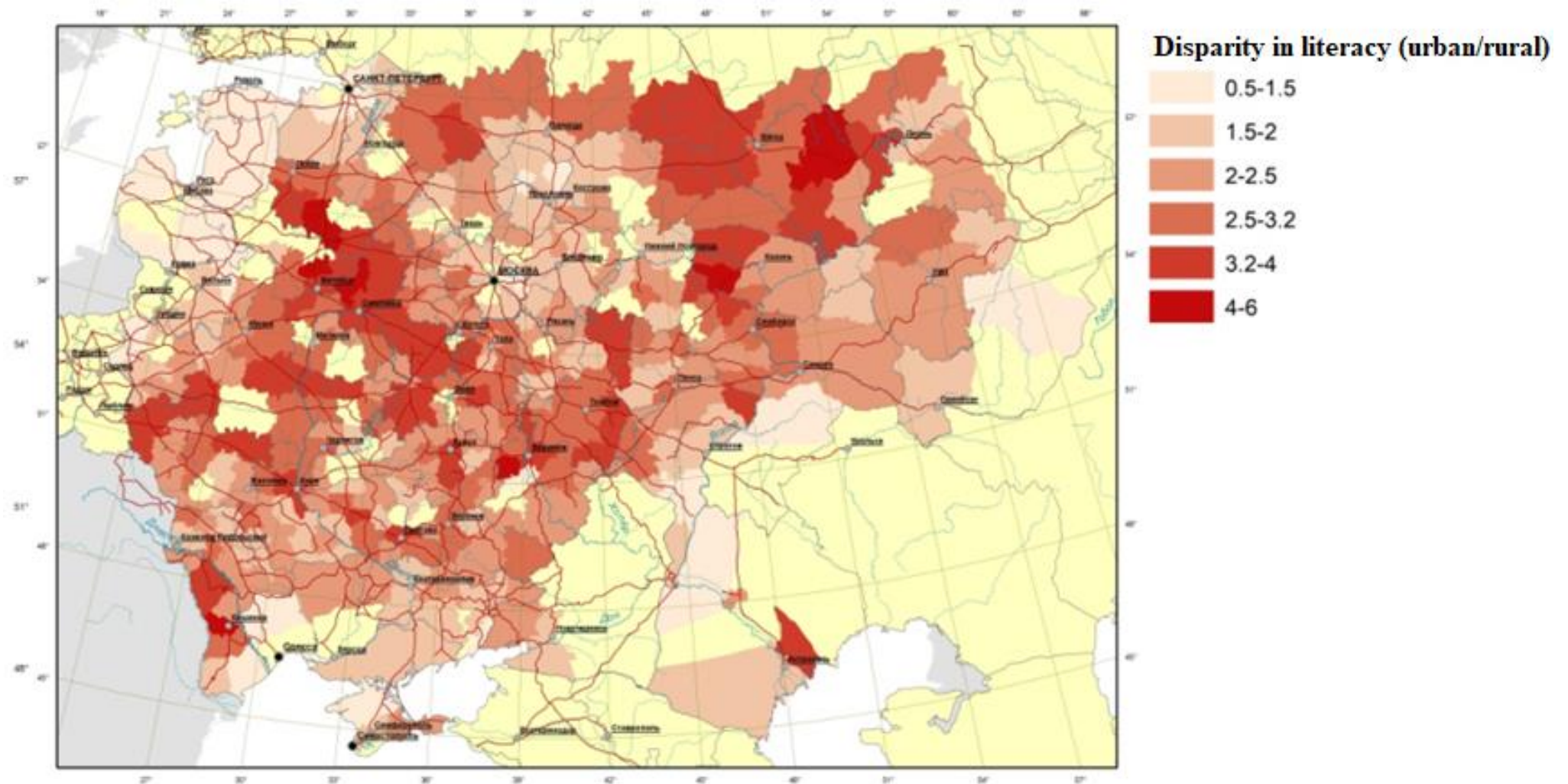


Fig. A2. Settlement disparity in the level of literacy of age groups of 20–29 and 40–49 year-olds in the uyezds of European Russia (average sample)

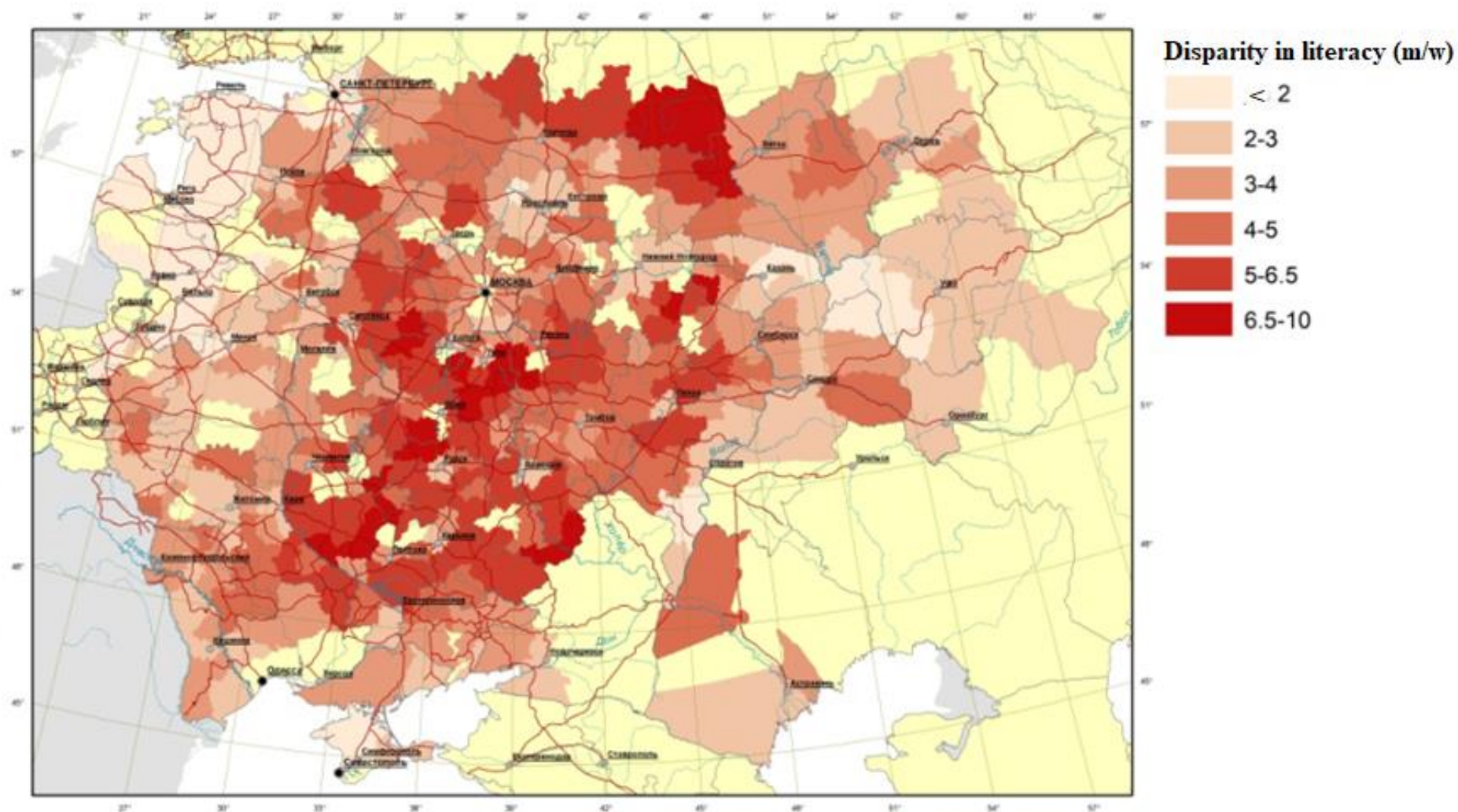


Fig. A3. Gender disparity in the level of literacy of age groups of 20–29 and 40–49 year-olds in the uyezds of European Russia (average sample)

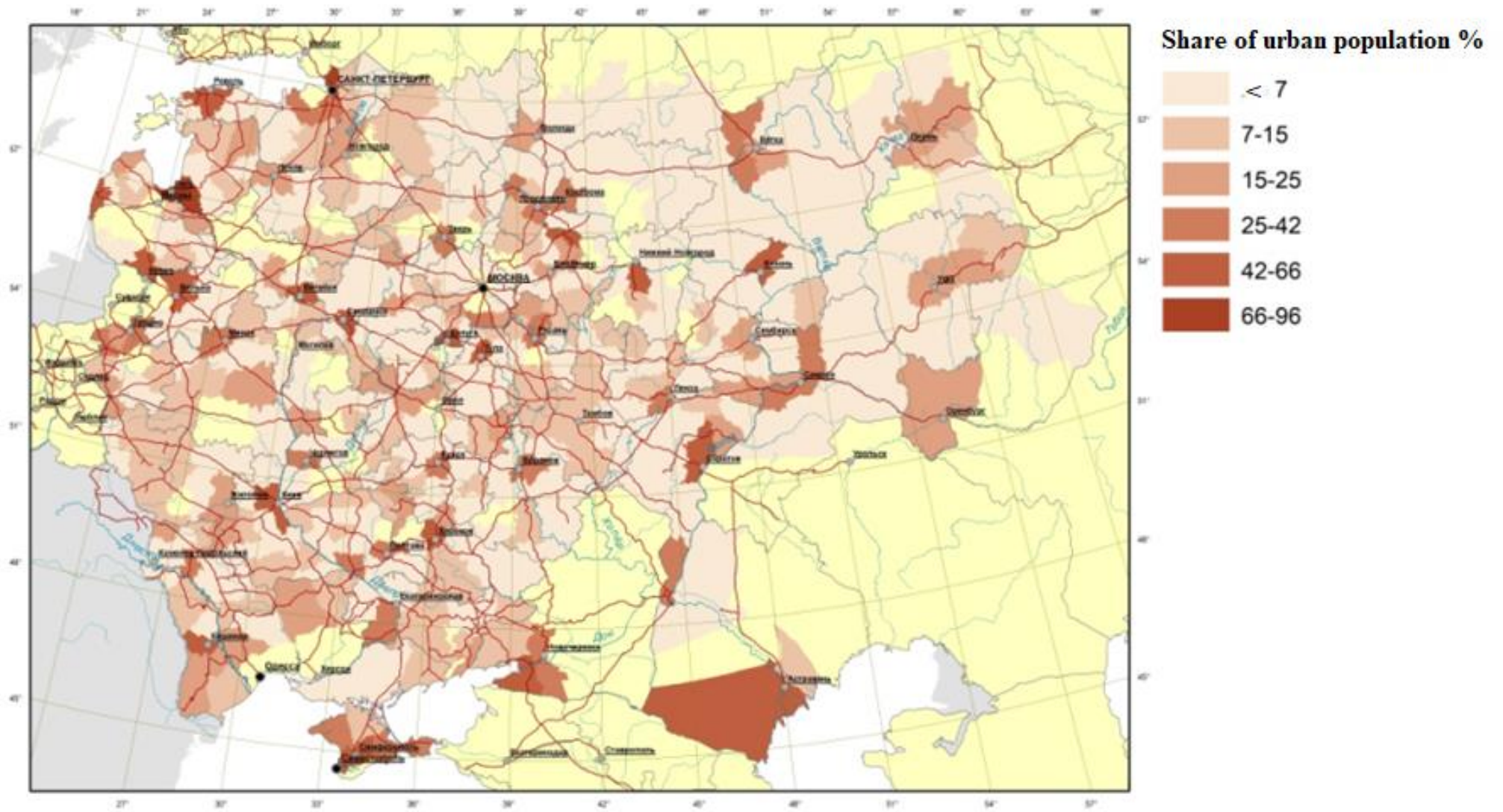


Fig. A4. Level of urbanization by uyezds of European Russia (average sample)

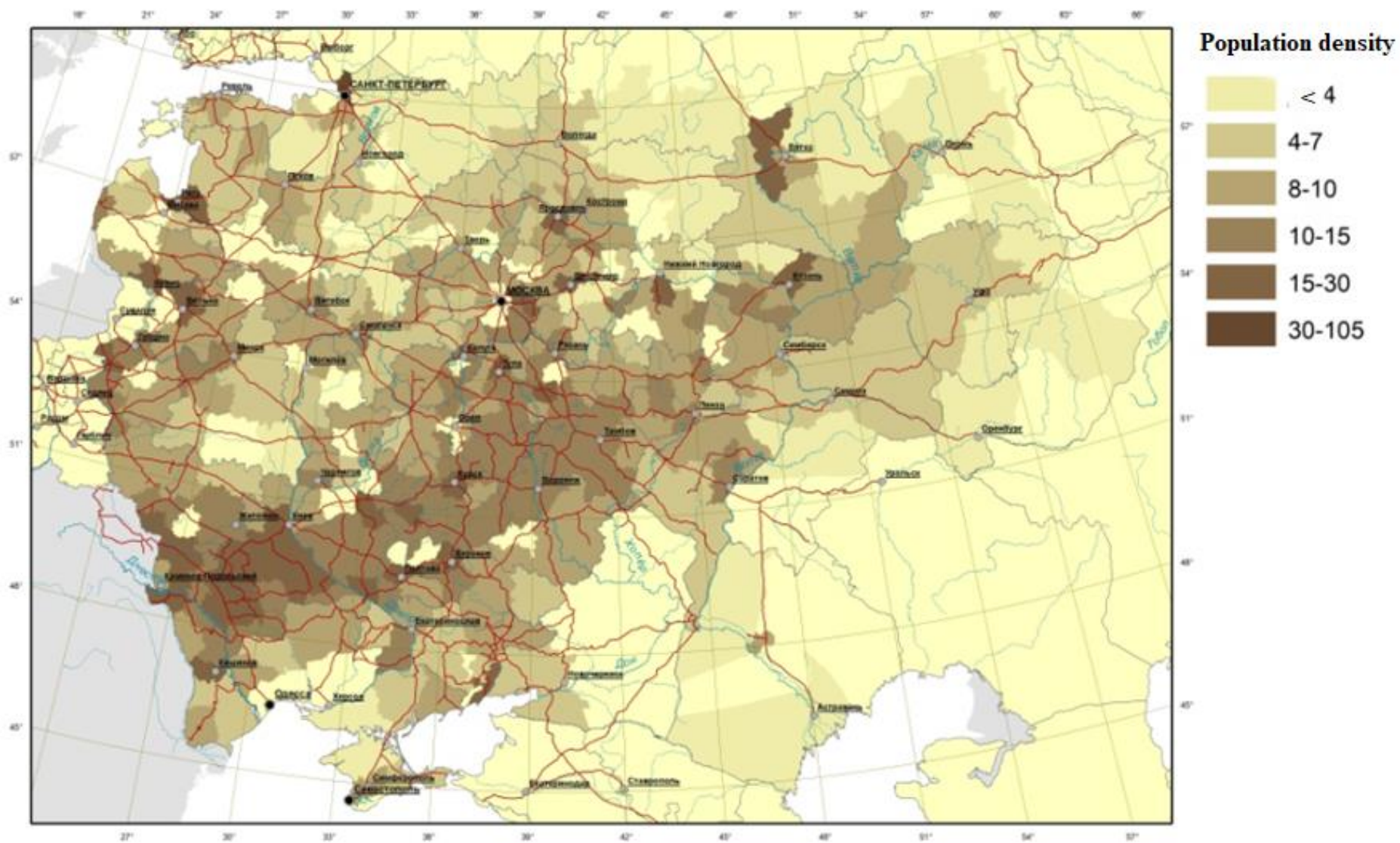


Fig. A5. Population density by uyezds of European Russia (average sample)



Fig. A6. Transport infrastructure, administrative borders, and centers of the governorates of European Russia

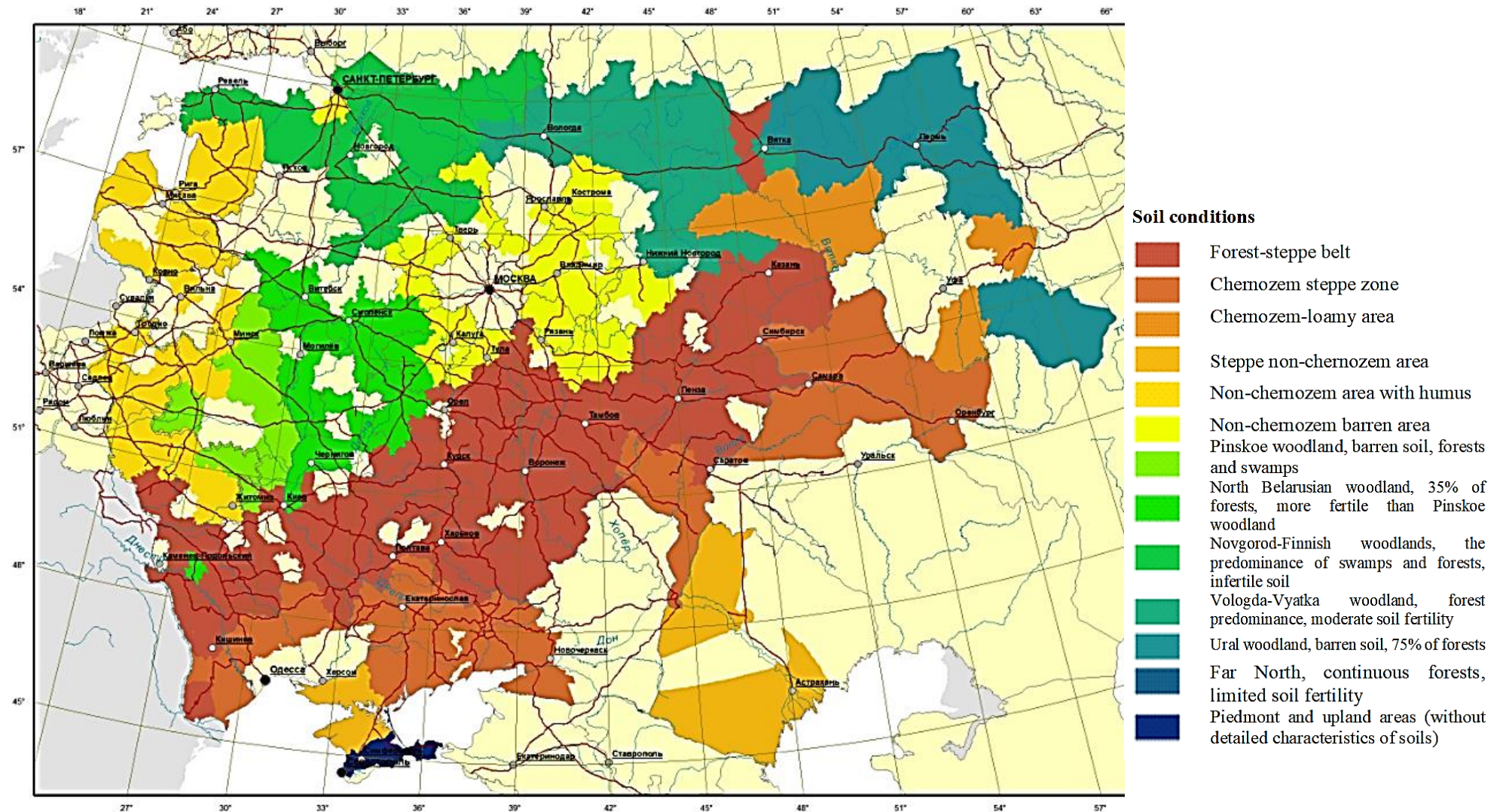


Fig. A7. Soil cover by uyezds of European Russia (average sample with truncation)

Список литературы

- Миров Л.В. Великорусский пахарь и особенности российского исторического процесса. Москва: РОССПЭН, 2001. 576 с.
- Миронов Б.Н. Российская империя: от традиции к модерну: в 3 т. Санкт-Петербург: ДМИТРИЙ БУЛАНИН, 2018. Т. 1. 896 с.
- Нуреев Р.М., Латов Ю.В. Экономическая история России (опыт институционального анализа). Москва: КНОРУС, 2016. 268 с.
- Пайнс Р. Россия при старом режиме / перевод с английского В. Козловского. Москва: Независимая газета, 1993. 424 с.
- Baten J., Hippe R. Geography, land inequality and regional numeracy in Europe in historical perspective // *Journal of Economic Growth*. 2018. Vol. 23, № 1. P. 79–109.
- Galor O., Moav O., Vollrath D. Land Inequality and the Emergence of Human Capital Promoting Institutions // *Review of Economic Studies*. 2009. Vol. 76, № 1. P. 143–179.
- Goldstone J.A. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press, 1991. 608 p.
- Engerman S.L., Sokoloff K.L. Factor Endowments, Inequality, and Paths of Development among New World Economies // *Economía Journal*. 2002. Vol. 3, № 1. P. 41–109.
- Sokoloff K.L., Engerman S.L. History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World // *Journal of Economic Perspectives*. 2000. Vol. 14, № 3. P. 217–232.
- Turchin P., Nefedov S.A. *Secular cycles*. Princeton, NJ: Princeton University Press, 2009. 360 p.

References

- Baten, J., and R. Hippe. “Geography, Land Inequality and Regional Numeracy in Europe in Historical Perspective.” *Journal of Economic Growth*, vol. 23, no. 1 (2018): 79–109.
- Engerman, S.L., and K.L. Sokoloff. “Factor Endowments, Inequality, and Paths of Development among New World Economies.” *Economía Journal*, vol. 3, no. 1 (2002): 41–109.
- Galor, O., O. Moav, and D. Vollrath. “Land Inequality and the Emergence of Human Capital Promoting Institutions.” *Review of Economic Studies*, vol. 76, no. 1 (2009): 143–79.
- Goldstone, J.A. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press, 1991.
- Milov, L.V. *Velikorusskii pakhar' i osobennosti rossiiskogo istoricheskogo protsessa* [The Great Russian ploughman and the peculiarities of the Russian historical process]. Moscow: ROSSPEN, 2001. (In Russian)
- Mironov, B.N. *Rossiiskaya imperiya: ot traditsii k modern* [The Russian Empire: from tradition to modernity]. 3 vols. Vol. 1. St Petersburg: DMITRII BULANIN, 2018. (In Russian)
- Nureev, R.M., and Yu.V. Latov. *Ekonomicheskaya istoriya Rossii (opyt institutsional'nogo analiza)* [Economic history of Russia (studies in institutional analysis)]. Moscow: KNORUS, 2016. (In Russian)
- Pipes, R. *Rossiya pri starom rezhime* [Russia under the old regime]. Translated from English by V. Kozlovskii. Moscow: Nezavisimaya gazeta, 1993. (In Russian)

Sokoloff K.L., and S.L. Engerman. “History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World.” *Journal of Economic Perspectives*, vol. 14, no. 3 (2000): 217–32.

Turchin P., and S.A. Nefedov. *Secular Cycles*. Princeton, NJ: Princeton University Press, 2009.

Информация об авторах

Дмитрий Валерьевич Диденко – доктор экономических наук, кандидат исторических наук, ведущий научный сотрудник Научно-исследовательской лаборатории экономической и социальной истории, профессор кафедры истории экономики, <http://orcid.org/0000-0001-5295-2538>, didenko-dv@ranepa.ru, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (д. 82, стр. 1, пр-т Вернадского, 119571 г. Москва, Россия).

Евгений Сергеевич Гришин – руководитель сектора исторической картографии и геоинформационных систем Научно-исследовательской лаборатории экономической и социальной истории, <https://orcid.org/0000-0002-9521-2246>, bibliosof-info@yandex.ru, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (д. 82, стр. 1, пр-т Вернадского, 119571 г. Москва, Россия).

Information about the authors

Dmitrii V. Didenko – Doctor of Science in Economics, Candidate of Science in History, Leading Researcher of the Research Laboratory of Economic and Social History, Professor of the Department of Economic History, <http://orcid.org/0000-0001-5295-2538>, didenko-dv@ranepa.ru, Russian Presidential Academy of National Economy and Public Administration (82, str. 1, pr. Vernadskogo, 119571 Moscow, Russia).

Evgenii S. Grishin – Head of the Department of Historical cartography and geo-information system of the Research Laboratory of Economic and Social History, <https://orcid.org/0000-0002-9521-2246>, bibliosof-info@yandex.ru, Russian Presidential Academy of National Economy and Public Administration (82, str. 1, pr. Vernadskogo, 119571 Moscow, Russia).

Заявленный вклад авторов

Определение цели и задач исследования на основе изучения научной литературы; обработка статистических данных; проведение расчетов; написание основной части текста – Д.В. Диденко (доля вклада автора – 2/3).

Определение круга источников; извлечение из источников (посредством цифровой обработки) данных о физико-географических, климатических, технологических показателях; написание части текста о соответствующих источниках, картографическая визуализация – Е.С. Гришин (доля вклада автора – 1/3).

Авторы заявляют об отсутствии конфликта интересов.

Author Contributions Statement

Determining the purpose and objectives of the study based on the study of research literature; processing of statistical data; carrying out calculations; writing the main part of the text by D. Didenko (the share of the author’s contribution is 2/3).

Determining the range of sources; extraction from sources (through digital processing) of data on physical-geographical, climatic, technological indicators; writing a part of the text about the

relevant sources, cartographic visualization by E. Grishin (the share of the author's contribution is 1/3).

The authors declare no conflict of interest.

Статья поступила в редакцию 28.01.2022; одобрена после рецензирования 11.04.2022; принята к публикации 19.04.2022.

The article was submitted 28.01.2022; approved after reviewing 11.04.2022; accepted for publication 19.04.2022.